

En esta edición

Trabajo decente
Encuestas complejas
Cuentas satélite del aire
Calidad estadística y muestreo
Distribución espacial de la vejez
Años de esperanza de vida perdidos
Análisis espacial de la mortalidad infantil
Diferenciales salariales por sindicalización
Estimación del tamaño de áreas pequeñas
Metodologías internacionales de retropolación
Flujos migratorios internacionales en Colombia
Análisis espacial de la pobreza multidimensional
Duración, probabilidad e incidencia del desempleo



Directivas del Departamento Administrativo Nacional de Estadística

Director

Mauricio Perfetti Del Corral

Subdirector

Diego Silva Ardila

Secretaria General

Alexandra Navarro Erazo

Dirección de Metodología y Producción Estadística

Eduardo Efraín Freire Delgado

Dirección de Síntesis y Cuentas Nacionales

Ana Victoria Vega Acevedo

Dirección de Geoestadística

Miguel Ángel Cárdenas Contreras

Dirección de Censos y Demografía

Liliana Acevedo Arenas

Dirección de Regulación, Planeación, Estandarización y Normalización

Nelcy Araque García

Dirección de Difusión, Mercadeo y Cultura Estadística

Soraya Peñaranda Gómez (e)

Departamento Administrativo Nacional de Estadística

Centro Andino de Altos Estudios
CANDANE

Carrera 59 no. 26 – 70 Int 1 CAN

PBX: 5978300 Ext. 2399

Teléfono directo: 5738020

cavendanop@dane.gov.co



Qué es la Revista ib

La Revista ib es una publicación periódica, arbitrada por pares evaluadores, orientada a difundir trabajos originales enmarcados en actividades de investigación para crear un espacio de difusión de datos y metodologías asociados a la actividad estadística, económica y demográfica.

El centro nacional de formación en estadística del DANE, CANDANE, obedeciendo a su naturaleza de ente académico, publica la Revista ib para ofrecer un espacio de debate, crítica y análisis de temas estadísticos, económicos y demográficos de alto nivel académico.

A través de esta publicación el DANE responde a su compromiso con la difusión y socialización de estadísticas, como un valor agregado anexo a su deber de producción y procesamiento oficial, al difundir artículos relacionados con información básica en temas demográficos, económicos, sociales, geográficos, agropecuarios y ambientales, que sirvan como instrumento de análisis de todos aquellos que utilizan la información estadística como insumo de su quehacer diario.

Las opiniones expresadas en los artículos son responsabilidad exclusiva del autor, su contenido no compromete al Departamento Administrativo Nacional de Estadística, ni a la Revista ib de la Información Básica Estadística.

Cuando un artículo es aceptado para publicación, los derechos de copia del mismo pasan a ser propiedad del DANE.

ib Información Básica en Estadística

ISSN 2256-1552

Vol. 3 Núm. 1

236 páginas

Enero 2014 - diciembre 2014

Director CANDANE

Bernardo Guerrero Lozano

Coordinadora editorial

Carolina Avendaño Pabón

Equipo editorial

Manuel Alejandro Godoy Cubillos

Carolina Avendaño Pabón

Equipo Investigación CANDANE

Néstor Eduardo Muñoz Rojas

Manuel Alejandro Godoy Cubillos

Comité editorial

Bernardo Guerrero Lozano

(Colombia)

*Departamento Administrativo Nacional
de Estadística*

Beatriz Piedad Urdinola Contreras

(Colombia)

Universidad Nacional de Colombia

Francisco Azuero Zúñiga

(Colombia)

Universidad de Los Andes

Leonardo Trujillo

(Colombia)

Universidad Nacional de Colombia

Corrección de estilo

Luis Javier Sánchez Duque

Diseño

Jaime Alberto Bustos Salazar

Diagramación

Diana María Jara Rivera

Impresión

Imprenta Nacional de Colombia

Bogotá, D.C. - Colombia

Publicación anual

Tiraje: 1100 ejemplares

Revista ib versión digital

[http://www.dane.gov.co/files/publicaciones/
index/](http://www.dane.gov.co/files/publicaciones/index/)

Índice

Página

- 5** **Duración, probabilidad e incidencia del desempleo en Colombia** / *Duration, probability and incidence of unemployment in Colombia.*
Jaime Tenjo Galarza (Colombia)
Martha Misas Arango (Colombia)
Alejandro Gaviria Jaramillo (Colombia)
Alfredo Contreras Eitner (Colombia)
- 29** **Una revisión a los procesos de aseguramiento de la calidad en operaciones estadísticas por muestreo** / *A review of the processes of quality assurance in statistics operations*
Leonardo Trujillo Oyola (Colombia)
Luz Mery González García (Colombia)
- 41** **Diferencial salarial entre sindicalizados y no sindicalizados: un análisis para los sectores público y privado** / *Wage differential between union and non-union workers: an analysis for public and private sectors*
José Daniel Trujillo Suárez (Colombia)
- 55** **Distribución espacial de la vejez en Colombia Censo General 2005** / *Spatial distribution of old age in Colombia General Census 2005*
Angela María Jaramillo de Mendoza (Colombia)
- 69** **Los años de esperanza de vida perdidos en adultos mayores. Bogotá, Cali y Medellín (1999-2009)** / *Years of life expectancy lost in older adults. Bogota, Cali and Medellin (1999-2009)*
Óscar Augusto Castellanos Ospina (Colombia)
Doris Cardona Arango (Colombia)
- 84** **Análisis espacial multitemporal de la tasa de mortalidad infantil municipal en Colombia** / *Multitemporal spatial analysis of the rate of infant municipal mortality in Colombia*
Javier Sebastián Ruiz Santacruz (Colombia)
Carlos Alberto Durán Gil (Colombia)
- 105** **Ajuste de estimadores mediante la técnica de propensity score en encuestas complejas** / *Adjusting of estimators by propensity score technique in complex surveys*
Hugo Andrés Gutiérrez Rojas (Colombia)
Luisa Fernanda Rojas Niño (Colombia)

Fotografía de portada por Horia Varlan

Publicada originalmente en: <http://www.flickr.com/photos/horiavarlan/4332388370/>
Bajo licencia Creative Commons Atribución

Página

- 115** **Viajeros vs. migrantes: una nueva metodología para la estimación de los flujos migratorios internacionales de Colombia** / *Travelers vs. migrants: a new methodology for estimating international migration flows of Colombia*
Joaquín Recaño Valverde (España)
Carolina Sánchez Barriga (Colombia)
Vivian Natalia Rivera Sepúlveda (Colombia)
- 136** **Compilación y síntesis de las metodologías internacionales aplicadas a procedimientos de retopolación** / *Compilation and synthesis of international methodologies applied to splicing procedures*
Sonia Milena Cifuentes Cruz (Colombia)
Andrés David García Ávila (q.e.p.d.) (Colombia)
- 161** **Cuenta de emisiones al aire: estudio de caso emisión de metano por la fermentación entérica del ganado bovino** / *Account air emissions: a case study of methane emission from bovine cattle enteric fermentation*
Luz Dary Yepes Rubiano (Colombia)
- 176** **Una comparación empírica entre algunos métodos demográficos y estadísticos para la estimación del tamaño de un área pequeña** / *An empirical comparison between some demographic and statistical methods for estimating the size of a small area*
Hugo Andrés Gutiérrez Rojas (Colombia)
José Fernando Zea Castro (Colombia)
- 189** **Un análisis del trabajo decente como instrumento conceptual** / *An analysis of decent work as a conceptual tool*
Claudia Carolina Córdoba Currea (Colombia)
- 205** **Análisis espacial de la pobreza multidimensional en Colombia a partir del censo de población de 2005** / *Spatial analysis of multidimensional poverty in Colombia from 2005 census*
Laura Estrada Arbeláez (Colombia)
Sandra Liliana Moreno Mayorga (Colombia)
- 229** **Submission of articles**
- 233** **Instrucciones para el envío de artículos**

Esta edición de la revista inicia con el trabajo realizado por Jaime Tenjo, Director del Programa de Economía de la Universidad Jorge Tadeo Lozano, y Martha Misas, Alfredo Contreras y Alejandro Gaviña, profesores investigadores asociados al Departamento de Economía de la Pontificia Universidad Javeriana. En esta investigación basada en la Encuesta Nacional de Hogares, los autores contribuyen al entendimiento de la dinámica laboral al investigar sobre la probabilidad de que una persona se encuentre desempleada, la duración del desempleo y los factores que lo determinan. De esta forma, se proporciona información detallada sobre las características de los procesos de búsqueda de empleo y desempleo en nuestro país, aspectos relevantes para el buen diseño de políticas públicas de mercado laboral.

Más adelante, Claudia Carolina Córdoba, de la Dirección de Metodología y Producción Estadística del DANE, nos presenta un balance conceptual de los enfoques de empleos atípicos, trabajos precarios, trabajos informales y empleos de mala calidad, los cuales se comparan posteriormente con el concepto de trabajo decente.

Sumándose a la línea de mercado laboral, José Daniel Trujillo, también parte del equipo de la Dirección de Metodología y Producción Estadística de la entidad, estima en su artículo el efecto que tienen los sindicatos sobre el salario de los trabajadores, tanto públicos como privados. De hecho, el autor encuentra un efecto diferencial en el salario del sindicato, tanto para el total, como para el sector privado y público.

Esta edición de la revista incluye también dos artículos que toman particular relevancia debido a los desafíos que trae consigo el envejecimiento demográfico que está enfrentando el mundo en la actualidad. Por un lado, Ángela María Jaramillo, docente investigadora de la Pontificia Universidad Javeriana, nos muestra de una forma clara la relación entre el proceso general de poblamiento en Colombia en el siglo XX y la distribución actual de la vejez, describiendo de una forma muy detallada las principales características de este grupo poblacional, de acuerdo con su distribución urbano-rural, lugar de nacimiento, hogares particulares, lugares especiales de alojamiento y el tipo de vivienda.

En este mismo sentido, y teniendo en cuenta que la transición demográfica conlleva a que la morbilidad y la mortalidad presenten cambios, pasando de una prevalencia de enfermedades transmisibles y parasitarias a una mayor prevalencia de enfermedades crónicas y degenerativas, Óscar Augusto Castellanos, de la Dirección de Censos y Demografía del DANE, y Doris Cardona, docente de la Facultad de Medicina de la Universidad CES (Comprometidos con la Excelencia) de Medellín, describen de una forma minuciosa el perfil de mortalidad de los adultos mayores en Bogotá, Medellín y Cali entre 1999 y 2009. Este ejercicio nos muestra la necesidad de contar con buenas fuentes de información, que permitan el direccionamiento de los programas de salud.

Por su parte, Javier Sebastián Ruiz, de la Dirección de Censos y Demografía de la entidad, y Carlos Alberto Durán, vinculado al área de Geoestadística del DANE, abordan el tema de infancia a través de un artículo en el que describen concienzudamente el desarrollo del análisis espacial de la Tasa de Mortalidad Infantil municipal en Colombia obtenida mediante las Estadísticas Vitales del DANE para 2004, 2007 y 2010, y encuentran una autocorrelación espacial diferencial a través del tiempo.

Joaquín Recaño, investigador de la Universidad Autónoma de Barcelona, en equipo con Carolina Sánchez, de nuestra Dirección de Censos y Demografía, y Vivian Rivera, de la Universidad Antonio Nariño, aplican una sugestiva metodología para presentarnos una estimación de flujos migratorios internacionales de Colombia. Esta investigación toma particular relevancia debido a que la medición de la migración

internacional es uno de los puntos más débiles de la información demográfica a nivel nacional.

A nivel de estadísticas relacionadas con medio ambiente, esta edición incluye un valioso ejercicio piloto de cuantificación de la emisión de gases de efecto invernadero, mediante el cual se proporciona elementos conceptuales para entender y dimensionar el desarrollo de la contabilidad ambiental económica. Asimismo, Luz Dary Yepes, su autora, vinculada a la Dirección de Cuentas Nacionales del DANE, nos muestra la necesidad de estructurar la cuenta de emisiones de aire como una herramienta de política enfocada al cambio climático.

La pobreza es también un tema de uno de los artículos de la revista, desarrollado por dos profesionales del DANE: Laura Estrada, de la Dirección de Metodología y Producción Estadística, en equipo con Sandra Liliana Moreno, del área de Geoestadística, exploran la dimensión espacial de la pobreza multidimensional a nivel municipal de nuestro país.

Teniendo en cuenta que en la actualidad existe un crecimiento del interés generalizado por contar con series de tiempo históricas consistentes, Sonia Milena Cifuentes, de la Dirección de Síntesis y Cuentas Nacionales del DANE, presenta los resultados de un trabajo realizado en equipo con Andrés García q.e.p.d., a través de los cuales se compendian las experiencias de los institutos de estadística internacionales más sobresalientes en el empalme de series de tiempo, con el objetivo proporcionar información adecuada y suficiente para dar a conocer métodos que son apropiados para garantizar la continuidad en la información publicada.

Finalmente, a nivel estadístico Leonardo Trujillo y Luz Mery González, profesores del Departamento de Estadística de la Universidad Nacional, presentan una detallada descripción histórica de las operaciones estadísticas por muestreo estadístico, además de una cuidadosa revisión de los lineamientos que se deben tener en cuenta para el aseguramiento de la calidad de las estadísticas a nivel internacional. Este artículo aporta valiosas recomendaciones a los lineamientos del proceso actual de aseguramiento de la calidad en dichas operaciones en nuestro país.

Hugo Andrés Gutiérrez, decano de la Facultad de Estadística de la Universidad Santo Tomás, acompañado de José Fernando Zea, profesor investigador de la misma facultad, realizan una comparación empírica entre algunos métodos demográficos y estadísticos para la estimación del tamaño de áreas pequeñas.

Para finalizar, respecto al tema estadístico, Luisa Fernanda Rojas, estadística de Nielsen Colombia, y Hugo Andrés Gutiérrez nos aportan un artículo conducente a enfrentar el sesgo que genera la ausencia de respuestas en los estimadores de totales de interés.

Con este número el DANE, al celebrar sus 60 años de existencia, presenta a los lectores de la presente edición de la revista nuevas contribuciones a la generación de conocimiento basado en las estadísticas oficiales.

Bernardo Guerrero Lozano
Director CANDANE

Duración, probabilidad e incidencia del desempleo en Colombia

Jaime Tenjo Galarza

Martha Misas Arango

Alejandro Gaviria Jaramillo

Alfredo Contreras Eitner

Duración, probabilidad e incidencia del desempleo en Colombia

Jaime Tenjo Galarza: Economista de la Universidad Nacional de Colombia, magíster en Economía y Doctorado en Economía de la Universidad de Toronto, Canadá. Director del Departamento de Economía de la Universidad Jorge Tadeo Lozano.

Correo electrónico: jaime.tenjog@utadeo.edu.co

Martha Misas Arango: Matemática de la Universidad Nacional de Colombia, magíster en estadística aplicada de la Universidad de Michigan, Estados Unidos. Profesora investigadora del Departamento de Economía de la Pontificia Universidad Javeriana.

Correo electrónico: mmisas@javeriana.edu.co

Alejandro Gaviria Jaramillo: Economista y magíster en Economía de la Universidad Javeriana. Profesor investigador del Departamento de Economía de la Pontificia Universidad Javeriana.

Correo electrónico: agaviria@javeriana.edu.co

Alfredo Contreras Eitner: Economista y magíster en economía de la de la Pontificia Universidad Javeriana. Director del Pregrado en Economía de la Universidad del Rosario.

Correo electrónico: alfredo.contreras@urosario.edu.co

Fecha de recepción: 26 de marzo de 2013

Fecha de aceptación: 6 de agosto de 2013

Resumen: la tasa de desempleo de una economía se considera el mejor indicador sintético de lo que sucede en el mercado laboral. Sin embargo, a pesar de sus ventajas, dicho indicador no arroja información sobre muchos aspectos de lo que sucede en dicho mercado. Una forma de aproximarse al entendimiento de esta dinámica es investigando tanto la probabilidad de que una persona se encuentre desempleada, como la duración del desempleo y los factores que la determinan. Estos son temas ampliamente estudiados en la literatura sobre economía laboral, pero que en el caso colombiano no han recibido la atención que ameritan. Este artículo se centra en la estimación de la duración del desempleo y la probabilidad de estar en desempleo para Colombia. El objetivo fundamental es proporcionar información detallada sobre las características de los procesos de búsqueda de empleo y del desempleo en Colombia, incluyendo la estimación de la incidencia del desempleo. Este trabajo busca proporcionar información útil para el desarrollo de políticas de mercado laboral.

Palabras clave: tasa de desempleo, mercado laboral, duración del desempleo, incidencia del desempleo.

Abstract: the unemployment rate in an economy is considered the best synthetic indicator of what happens in the labor market. However, despite its advantages, this indicator does not yield information on several matters of what occurs in that market. A first approach to the dynamics of unemployment is to investigate both the probability of being unemployed, and the duration of unemployment considering its determinants. These are topics widely studied by the labor economics literature except in Colombia where it is an almost unexplored issue. This article focuses on the estimation of both the duration of unemployment and the probability of being unemployed for Colombia. The main objective is to provide detailed information of the processes of job search, unemployment and its characteristics, including the estimation of the incidence of unemployment. This paper aims to provide useful information for designing new policies for the labor market.

Key words: unemployment rate, labor market, unemployment duration, unemployment incidence.

1. Introducción

En general la tasa de desempleo de una economía se considera el mejor indicador sintético de lo que sucede en el mercado laboral. Sin embargo, a pesar de sus ventajas, dicho indicador no arroja información sobre muchos aspectos de lo que sucede en dicho mercado. Una de estas fallencias tiene que ver con el hecho de que una misma tasa de desempleo puede generarse bajo diferentes condiciones y un cambio en su nivel se puede deber a cambios en cualquiera de las condiciones que concurren a su determinación. Más específicamente, una determinada tasa de desempleo se puede deber a una situación en la cual los miembros de la fuerza de trabajo quedan desempleados con mucha frecuencia (alta incidencia), pero duran desempleados poco tiempo (baja duración); o a que quedan desempleados con baja frecuencia pero duran buscando empleo mucho tiempo. (Por ejemplo, una tasa mensual de desempleo de 12% se puede deber a que 1% de los trabajadores quedan desempleados cada mes, y duran buscando empleo 12 meses, o a que 6% de los trabajadores quedan desempleados cada mes, pero consiguen un empleo nuevo en 2 meses). Distinguir entre estos dos tipos de situaciones es importante tanto para el diseño de políticas laborales adecuadas, como para lograr un mejor entendimiento del funcionamiento del mercado laboral.

Una forma de aproximarse al entendimiento de esta dinámica es investigando tanto la probabilidad de que una persona se encuentre desempleada, como la duración del desempleo y los factores que la determinan. Estos son temas ampliamente estudiados en la literatura sobre economía laboral, pero que en el caso colombiano no han recibido la atención que ameritan.

Este artículo se centra en la estimación de la duración del desempleo y la probabilidad de estar en desempleo para Colombia. El objetivo fundamental es proporcionar información detallada sobre las características de los procesos de búsqueda de empleo y del desempleo en Colombia, incluyendo la estimación de la *incidencia* del desempleo, un elemento relevante para el diseño de políticas de mercado laboral. La estimación de los modelos se basa en la información contenida en la Gran Encuesta Integrada de Hogares para cabecezas municipales correspondiente al tercer trimestre (julio, agosto y septiembre) de 2010.

El artículo está dividido en ocho secciones. La primera es esta introducción. En la segunda se presenta la revisión de literatura sobre modelación de la duración del desempleo tanto a nivel internacional como a nivel nacional. En la tercera sección se presenta un modelo simple de búsqueda de empleo para dar sustento teórico a las estimaciones empíricas de las siguientes secciones. La cuarta se centra en el estimador de Kaplan-Meier, como medida no paramétrica de la duración del desempleo y sus resultados para el caso colombiano. En la quinta se consignan las metodologías econométricas de duración y probabilidad. La sexta describe la información utilizada en los modelos de duración y probabilidad; en particular, la construcción de las variables asociadas a las condiciones del mercado referentes a duración del desempleo y salarios por oficio. En la siguiente sección, se presentan los resultados de los modelos de duración y probabilidad. Finalmente, en la última sección se muestran los resultados de las diferentes duraciones del desempleo por perfiles de individuo y tipo modelo y la incidencia del desempleo en Colombia.

2. Revisión de la literatura (con énfasis en métodos de estimación)

Uno de los primeros trabajos sobre la duración del desempleo es el de Ehrenberg y Oaxaca (1976); en este se realiza una primera aproximación a la explicación de la duración del desempleo. Los datos que se utilizan en el estudio provienen del National Longitudinal Survey (NLS). Este trabajo introduce a la explicación de la duración del desempleo los fundamentos de las teorías de búsqueda de empleo y enfatiza en la importancia de controlar por la variable edad.

En 1979 se publica el primer artículo que explica cómo se debe modelar la duración del desempleo superando los problemas de censura y heterogeneidad presentados en trabajos anteriores. Lancaster (1979) estudia los diferentes problemas que se presentan cuando se interpreta el efecto de las diferencias entre los individuos que buscan empleo y el tiempo que tardan en conseguirlo. El estudio se desarrolla basándose en las teorías de búsqueda de empleo. Los datos utilizados son el resultado de encuestas realizadas sobre una muestra de trabajadores británicos no calificados. Adicionalmente, Lancaster define una lista de variables que deben ser consideradas en la explicación de la duración del desempleo. Estas variables son:

- i) Tiempo desempleado: Esta variable se refiere al tiempo transcurrido entre el momento en que el individuo perdió el último empleo y el momento de la encuesta.
- ii) Empleo reciente: La pregunta se refiere a si el individuo ha

encontrado empleo recientemente; es necesario, por tanto, tener observaciones de individuos que hayan terminado su período de desempleo dentro de la muestra. Es decir, se tienen en cuenta aquellos individuos que estaban en desempleo pero que durante el período de estudio encontraron empleo.

iii) Edad: Esta variable permite explicar muchas diferencias entre los individuos, tales como el número de ofertas recibidas.

iv) Desempleo: Definido como el porcentaje de la población que está desempleada en la zona donde vive el individuo entrevistado. Tal como lo sugiere Lancaster, esta variable recoge dos efectos muy cercanos: probabilidad de conseguir empleo y expectativas de conseguirlo. En la sección V se presentará una nueva propuesta para medir expectativas de desempleo.

v) Reemplazo: Esta variable mide todos los otros ingresos que puede tener una persona, determinantes fundamentales para que el individuo realice un mayor o menor esfuerzo en su búsqueda de empleo. Esta variable es proporcional al salario de reserva de los individuos.

Meyer (1990) muestra, tanto de manera teórica como empírica, que el seguro de desempleo tiene un efecto positivo sobre la duración de este (un mayor seguro de desempleo conlleva una mayor duración del desempleo). Adicionalmente, introduce el uso de análisis semiparamétrico para construir indicadores que permitan diagnosticar la existencia de variables omitidas. La muestra es conformada por hombres en 12 estados para el período comprendido entre 1978 y 1983. La fuente es Continuous Wage and Benefit History (CWBH).

Pollman-Schult y Büchel (2005) siguen la línea de Meyer (1990), pero no solo se refieren al seguro de desempleo sino también a una medida más amplia que denominan beneficios del desempleo. Los autores sostienen que mayores beneficios del desempleo aumentarán la duración del mismo, pero mejorarán el *match* que se obtenga al final del período de desempleo. A través del estimador Kaplan-Meier y de modelos de regresión Hazard llegan a la conclusión de que las personas que no reciben beneficios del desempleo efectivamente tienen menor duración de este pero también tienen mayores tasas de transición a sobre-educarse. Adicionalmente, encuentran que el hecho de recibir beneficios de desempleo no explica la transición a un empleo más calificado. Los datos son tomados del West German Life History Study, recolectado en 1998 para cohortes de individuos nacidos entre 1964 y 1971.

Los trabajos más recientes han desarrollado metodologías más convenientes para el análisis de duración del desempleo. Por ejemplo Fitzenberger y Wilke (2005) desarrollan el método de regresión cuantílica con datos censurados para análisis de duración, con aplicaciones a la duración del desempleo; su aplicación se centra en información del oeste de Alemania. Sin embargo, este método aún no permite incluir covariables variantes en el tiempo y no presenta solución a posibles problemas de heterogeneidad no observada. Es de señalar que este trabajo retoma los desarrollos de Koenker y Billias (2001).

En general, un problema que se enfrenta al trabajar datos de duración del desempleo provenientes de encuestas es que la no respuesta a estas puede causar diferentes tipos de sesgo en los resultados obtenidos. Van der Berg y Dolton (2006) desarrollan una metodología para identificar entre dos

posibles explicaciones del sesgo en los resultados: el primero de selectividad por determinantes no observados de la duración del desempleo y el segundo un efecto causal del motivo de salida del último trabajo y, por consiguiente, la no respuesta de la encuesta.

De acuerdo con Wichert y Wilke (2008), los estimadores más utilizados no necesariamente son aplicables dada la naturaleza del proceso generador de los datos recolectados y pueden causar diferentes formas de censura. Para solucionar este problema se propone una serie de estimadores no paramétricos tipo Kaplan-Meier para las tasas Hazard condicionales y para la función cuantílica condicional. Con este nuevo estimador se superan los problemas que se presentan al emplear los estimadores tradicionales.

Dentro de los trabajos más relevantes para Colombia está el de Tenjo y Ribero (1998) donde además de la duración del desempleo se trabaja la participación laboral. Este estudio utiliza modelación bajo distribución Weibull y realiza estimaciones por género, controlando por estado civil. Emplea datos de la Encuesta Nacional de Hogares para el mes de junio, para los años 1988, 1994 y 1996. Encuentra una relación directamente proporcional entre el ingreso familiar y la duración del desempleo de las personas solteras e inversa con la duración en el caso de las personas casadas. Núñez y Bernal (1998) incluyen una sección dedicada a la duración del desempleo en el que utilizan datos del mes de junio de 1988, 1992 y 1996 de la Encuesta Nacional de Hogares. Martínez (2003) hace un estudio sobre duración de empleo y desempleo con datos de la Encuesta Nacional de Hogares.

Castellar y Uribe (2003) llevan a cabo un estudio para Colombia utilizando también modelos para-

métricos de duración tipo Weibull. Emplean datos de la Encuesta Nacional de Hogares correspondientes al segundo trimestre de los años 1988, 1992, 1994, 1996 y 1998. El estudio encuentra que la duración promedio del desempleo es de 42 semanas, considerando como determinantes de la duración del desempleo las variables ser jefe de hogar (relación negativa con la duración), ingresos no laborales (a menores ingresos no laborales hay una menor duración del desempleo), género (los hombres tienen menor duración), educación (efecto negativo y no lineal sobre la duración del desempleo), experiencia y dispersión salarial.

Viáfara y Uribe (2008) realizan un análisis no paramétrico para Colombia empleando el estimador Kaplan-Meier. Este trabajo se centra en los canales de búsqueda. Adicionalmente, presentan una amplia revisión de la literatura en cuanto a modelos de duración y estudios de canales de búsqueda para el caso colombiano.

3. Un modelo simple de búsqueda de empleo

El problema de búsqueda de empleo ha sido extensamente estudiado en la literatura económica. En dicha literatura se reconoce que la información no es un bien gratuito y que es rentable, desde el punto de vista privado¹, dedicar recursos a adquirir información en vez de aceptar el primer trabajo disponible. Una estrategia de búsqueda consiste en fijar un salario de reserva, por debajo del cual se rechaza cualquier oferta de trabajo.

¹ En general se reconoce que hay fallas de mercado que hacen que procesos de búsqueda (inversión en información) sin la adecuada infraestructura institucional no sean socialmente óptimos. Esto se debe a que debido a las externalidades existentes (búsquedas más prolongadas permiten una mejor asignación de recursos, menor rotación de la fuerza de trabajo y una mayor productividad) y al hecho de que la información, una vez obtenida, se convierte en un bien público, las decisiones privadas generalmente resultan en niveles de inversión inferiores a lo socialmente óptimo.

Dicho salario de reserva se determina de tal manera que maximice los beneficios netos futuros privados del proceso de búsqueda. Un resumen sencillo de lo que sería la estrategia es el siguiente.

Sea $g(w)$ la función de densidad que describe la percepción del individuo de lo que es la distribución de ofertas salariales en un momento determinado. $g: (\mu, \sigma)$.

Para su búsqueda de empleo, el trabajador fija un salario de reserva w_r tal que

Si $w \geq w_r$ acepta la oferta
Si $w < w_r$ rechaza la oferta

Dado lo anterior, la probabilidad de que una oferta sea rechazada estaría dada por:

$$Prob(w < w_r) = G(w_r) = \int_0^{w_r} g(w)dw \quad (1)$$

La probabilidad de que la oferta sea aceptada estaría dada por:

$$Prob(w \geq w_r) = 1 - G(w_r) \quad (2)$$

Si las ofertas son un proceso aleatorio, el valor esperado del número esperado de ofertas que se recibirían antes de que una sea aceptable está dado por:

$$D = E(L) = \frac{1}{1 - G(w_r)} \quad (3)$$

Si se hace el supuesto de que el trabajador recibe una oferta por período, D se puede interpretar como la duración esperada de la búsqueda. Si este supuesto no se cumple, entonces la duración de la búsqueda sería una función de D .

El salario esperado w_e que obtendría el trabajador una vez se reciba una oferta aceptable se puede expresar de la siguiente forma:

$$w_e = E(w|w \geq w_r) = \frac{\int_{w_r}^{\infty} wg(w)dw}{1 - G(w_r)} \quad (4)$$

Determinación del salario de reserva (w_r). Una forma sencilla de

plantear el problema es suponer que el trabajador fija su salario de reserva de tal manera que maximiza los *beneficios netos* futuros de la búsqueda (diferencia entre beneficios y costos), medidos en términos de valor presente.

Los procesos de búsqueda tienen costos directos y de oportunidad. Los costos directos son los asociados con la búsqueda misma, como la adquisición de información (consulta de avisos, bases de datos, etc.), la generación de ofertas (envío hojas de vida y otras cosas), etc. Los costos de oportunidad están asociados con los ingresos que deja de percibir el trabajador por dedicar su tiempo a la búsqueda en vez de trabajar (el salario que deja de percibir por período). En general, los costos directos se pueden considerar como fijos (por período), pero el costo de oportunidad crece con la duración de la búsqueda. Si suponemos que se recibe una oferta por período de tiempo, podemos expresar los costos (C) como una función de la duración misma de la búsqueda:

$$C = FD + h(D) = C(w_r) \text{ y } \frac{\partial C}{\partial w_r} > 0 \quad (5)$$

donde F representa el costo fijo directo por período y h es el costo de oportunidad (que es función de la duración de la búsqueda).

Los beneficios (Y) de la búsqueda están asociados con los ingresos que recibe el trabajador y se pueden medir como una función de w_e .

$$Y = Y(w_r); \text{ con } \frac{\partial Y}{\partial w_r} > 0 \text{ y } \frac{\partial^2 Y}{\partial w_r^2} < 0 \quad (6)$$

Lo anterior implica que el trabajador determina su salario de reserva óptimo maximizando el valor presente del beneficio neto ($V(B)$) de la búsqueda, dado por la siguiente función:

$$V(B) = \int_D^N Y(w_r)e^{-rt} dt - \int_0^D C(w_r)e^{-rt} dt \quad (7)$$

$$V(B) = \frac{1}{r} [Y(w_r)(e^{-rD} - e^{-rN}) - C(w_r)(1 - e^{-rD})]$$

El valor óptimo del salario de reserva (w_r^*) se obtiene resolviendo la siguiente ecuación para w_r :

$$\frac{\partial V(B)}{\partial w_r} = 0 \rightarrow w_r^* = k(g, r, N, X) \quad (8)$$

lo que es equivalente a igualar el beneficio marginal con el costo marginal de la búsqueda. La ecuación (8) indica que el salario de reserva óptimo es una función de la distribución de ofertas (g), la tasa de descuento, el horizonte para la toma de decisiones², N (tiempo relevante para efectos de la planea-

ción de la búsqueda y duración del empleo resultante) y un vector de otras variables X .

Una vez conocido w_r^* se puede obtener el valor de D reemplazando en la ecuación (3). De la misma manera, la duración esperada de la búsqueda sería:

$$D = d(g, r, N, X) \quad (9)$$

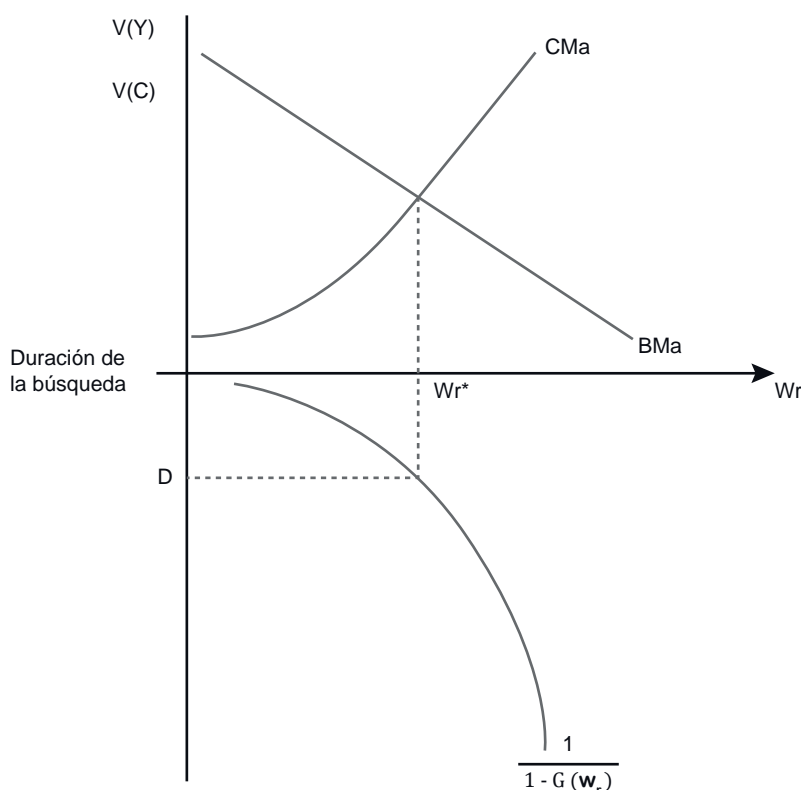
El gráfico 1 ilustra este procedimiento.

sencilla de incluir la diversidad de los mercados laborales. La desviación estándar (σ) se puede suponer constante, pero también se puede hacer depender de algunas características de las personas para captar diferencias en los mercados laborales. Determinar cuáles son las variables responsables por la variabilidad de σ puede ser un problema empírico.

b. El problema de la tasa de descuento. La tasa de descuento es una medida del costo intertemporal de los recursos necesarios para financiar la búsqueda de empleo. En situaciones donde los mercados funcionan adecuadamente, dicha tasa sería la tasa de interés de equilibrio de largo plazo. Sin embargo, en el caso de la financiación del desempleo el acceso a recursos financieros (en ausencia de intervención estatal) es muy limitado o nulo. Esto constituye una falla de mercado (mercados incompletos) y significa que la financiación de la búsqueda de empleo depende fundamentalmente de los recursos del individuo. Una forma de captar esto es haciendo la tasa de descuento una función inversa de la riqueza de los individuos o de los ingresos familiares. Esto llevaría a que (otras cosas iguales) la duración de la búsqueda de empleo aumente con los niveles de riqueza o de ingreso familiar y tendría la implicación de que para algunos (¿muchos?) individuos la inversión en información sobre mercado laboral será inferior a lo socialmente óptimo.

Por otro lado, se sabe que la efectividad de los procesos de búsqueda depende en gran medida de la efectividad de los canales de búsqueda y que entre estos, las redes de contacto personales (amigos, conocidos,

Gráfico 1. Nivel óptimo de salario de reserva.



Fuente: elaboración propia

La intersección entre el beneficio marginal de la búsqueda (Bma) y el costo marginal de la misma (Cma) determinan el nivel óptimo del salario de reserva (w_r^*). Una vez conocido dicho salario, se reemplaza en la función de duración (panel

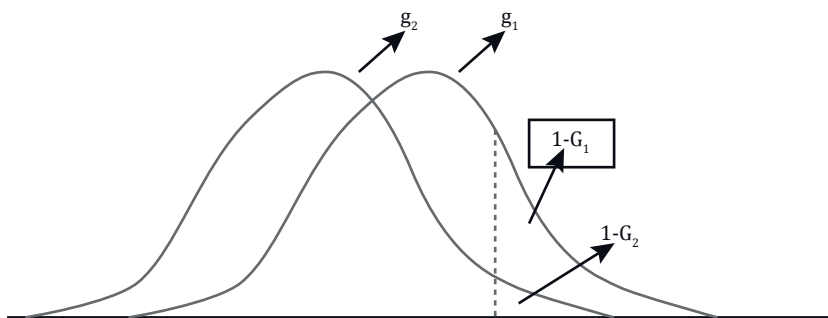
inferior del gráfico) para obtener la duración esperada.

Algunas extensiones:

a. μ se puede hacer función de variables de capital humano y otras que determinan los niveles de ingresos (por ejemplo, sexo, etc.). Esta es una forma

² En modelos más complejos el tiempo esperado de duración del empleo N puede hacerse una función de los resultados de la búsqueda.

Gráfico 2. Distribución de búsqueda de empleo.



Fuente: elaboración propia

etc.) son posiblemente los más importantes. Probablemente la calidad de estos contactos está relacionada con los niveles de riqueza o de ingreso familiar y, por lo tanto, las oportunidades de empleo son una función creciente de dicha riqueza.

De esta forma, el ingreso familiar juega dos papeles: por un lado representa las posibilidades de los individuos en el proceso de búsqueda y en ese sentido está asociado con las aspiraciones (nivel del salario de reserva); por el otro, es una proxy de la efectividad de los canales de búsqueda y en ese sentido está asociado con las oportunidades. Estos dos efectos, aspiraciones y oportunidades, operan en sentido contrario en la determinación de la duración de la búsqueda de empleo.

c. Equivocaciones en la distribución de ofertas salariales. La fijación del salario de reserva depende de la percepción de los individuos sobre lo que es la distribución de ofertas salariales en el mercado laboral. Sin embargo, dada su posición de trabajadores, la información disponible sobre las condiciones laborales fácilmente puede ser equivocada. Los trabajadores aprenden acerca del mercado

a través de los procesos de búsqueda de trabajo. Esto se puede ver más claramente de la siguiente manera:

Sea un trabajador que asume que la distribución de ofertas salariales es g_1 : (μ_1, σ) , cuando en realidad es g_2 : (μ_2, σ) , donde $\mu_1 > \mu_2$. Es decir, el trabajador sobrevalúa el verdadero valor de la media de la distribución de ofertas salariales. A partir de la distribución g_1 el trabajador escoge un salario de reserva w_{r1} , lo cual implica que en promedio la duración de la búsqueda sería:

$$D_1 = \frac{1}{1 - G_1}$$

Sin embargo, dado que la verdadera distribución es g_2 , la duración de la búsqueda sería:

$$D_2 = \frac{1}{1 - G_2} > D_1$$

En este caso el trabajador observa dos cosas: primero que la duración de su búsqueda se prolonga por más allá de lo esperado. Segundo, que después de un tiempo K la serie de ofertas recibidas $\{w_1, w_2, \dots, w_k\}$ no tiene la estructura distributiva que generaría la distribución g_1 . En general lo que observaría es que la proporción de ofertas laborales bajas tendría un peso mayor del que se debería esperar si dichas ofertas fueran generadas por g_1 .

En general el trabajador se ve obligado a revisar sus estimativos y trabajar con una distribución más cercana a g_2 , con el agravante de que ya ha usado parte de los recursos disponibles para la búsqueda de empleo.

Cuando el error es el contrario (subestimar la media de la distribución de ofertas), el trabajador encontrará un empleo aceptable más rápidamente de lo esperado. En este caso lo más probable es que el trabajador acepte el empleo pero no dure en él la cantidad de tiempo que esperaba.

4. Estimación no-paramétrica de la duración del desempleo

4.1. Estimador de Kaplan-Meier

La información utilizada para el ejercicio de estimación de la duración del desempleo, como ya se mencionó, proviene de la Gran Encuesta Integrada de Hogares, tercer trimestre de 2010 (agosto, septiembre y octubre). Dicha base de datos contiene información sobre las características personales como tiempo de desempleo, sexo, edad, educación, estado civil, posición en el hogar e ingresos del resto de la familia, todas ellas variables relevantes en la explicación de la duración del desempleo (ver revisión bibliográfica). Sin embargo, la variable duración (tiempo que llevan los desempleados buscando trabajo) es una variable censurada, en el sentido de que presenta el tiempo que llevan los trabajadores buscando empleo al momento de la encuesta, pero no se sabe cuánto tiempo más van a permanecer en esta condición.

Esta censura genera problemas porque introduce sesgos (subvalúa) en los indicadores sintéticos de duración del desempleo (como la media y la mediana). Nuestra revisión

de literatura a nivel nacional indica que ninguno de los estudios anteriores hechos en Colombia tienen en cuenta este problema y, por lo tanto, tampoco presentan alternativas de solución. Aún más, en muchas publicaciones oficiales y no oficiales en las que se usan medidas de duración de desempleo se presentan promedios, medianas y otras medidas de la duración del desempleo basadas en la información existente sin ningún tipo de corrección.

En este documento se propone una metodología para corregir el sesgo de censura, consistente en incluir en la base de datos el tiempo que les tomó conseguir el empleo a los trabajadores que han obtenido el empleo actual recientemente. Es decir, se conforma una variable que especifica qué individuos terminaron la búsqueda y cuáles de ellos continúan en esta. Esto permite tener información sobre la duración completa del desempleo y posibilita la construcción de estimativos consistentes sobre la duración de este.

Para la medición de la duración del desempleo se utilizaron modelos paramétricos y no paramétricos. Los no paramétricos incluyen la estimación de funciones de supervivencia a través del modelo de Kaplan y Meier, mientras que los paramétricos implican la estimación de modelos de duración ATF (Accelerated Time Failure) con el supuesto de que el error sigue una distribución Weibull.

Metodología no-paramétrica de la duración del desempleo. Una primera aproximación a la duración del desempleo se lleva a cabo a través de estadísticas no paramétricas y en particular mediante el estimador de Kaplan-Meier de la función de sobrevivencia (Hosmer, Lemeshow y May, 2008; Klein y Moeschberger, 2003; y Allison, 2004).

4.2. Resultados del análisis no paramétrico

A continuación se presentan los resultados del análisis de duración del desempleo considerando la Gran Encuesta Integrada de Hogares para los meses de julio, agosto y septiembre de 2010. La muestra objeto de estudio se conforma con base en la información de cada uno de estos meses, de tal forma que la variable censura se construye de la siguiente manera:

- La información de julio considera a los individuos no censurados como aquellos individuos que tienen un mes o menos de estar ocupados.
- La información de agosto considera a los individuos no censurados como aquellos individuos que tienen dos meses o menos de estar ocupados.
- La información de septiembre considera a los individuos no censurados como aquellos individuos que tienen tres meses o menos de estar ocupados.

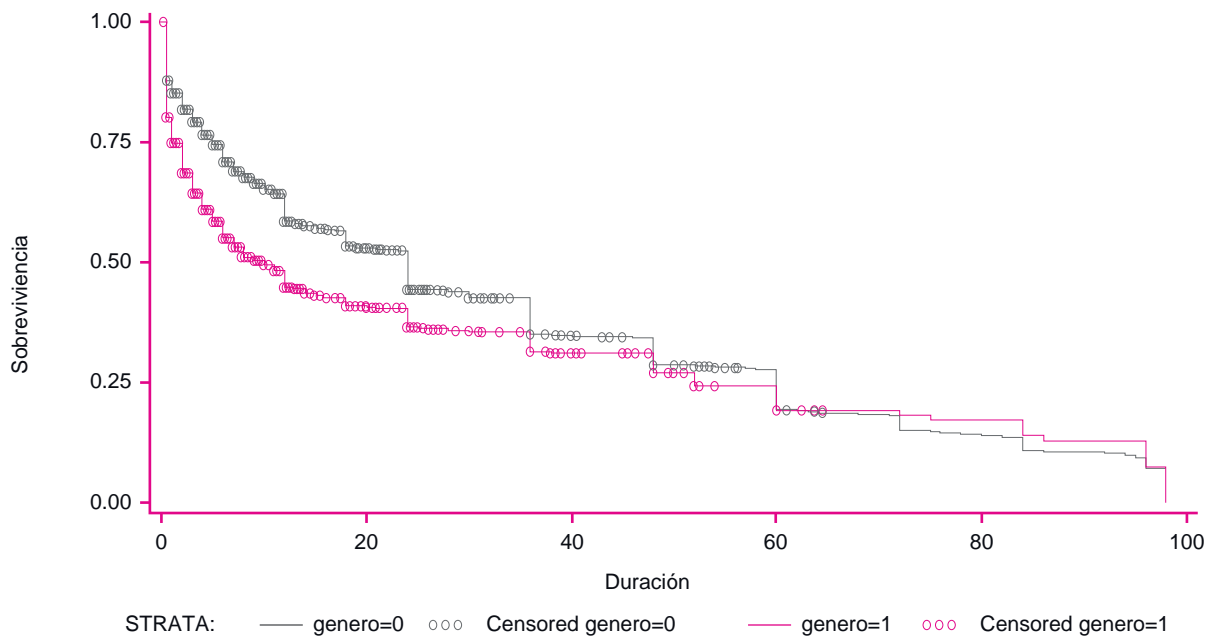
Así, la muestra objeto del trabajo reúne la información del tercer trimestre con la construcción de la variable censura bajo el esquema ya explicado. En este caso, la caracterización se realiza por la variable género, donde género = 1 si la observación corresponde a un hombre y 0 en caso de corresponder a una mujer. Como se deriva del gráfico 3, la duración del desempleo es menor para los hombres. En particular el 50% de los hombres han conseguido empleo en menos de un año, en tanto que, siendo mujer, el tiempo requerido para alcanzar el 50% es de 24 meses. La prueba estadística de Wilcoxon de diferencia de medias por género es igual a 0,0001, lo que muestra que el comportamiento entre los dos grupos es estadísticamente diferente.

Un análisis similar se hizo para estudiar las diferencias en la duración entre personas con distintas características: (i) jefe de hogar versus no-jefe de hogar, (ii) con o sin menores a cargo, (iii) por grupos de edad, (iv) con o sin mayores a cargo, (v) por grados aprobados de educación, (vi) por niveles de ingresos familiares y (vii) por estado civil. Los resultados de estos ejercicios aparecen resumidos en la tabla 1 considerando la muestra en su totalidad y las submuestras conformadas por hombres y mujeres.

i) La variable jefe de hogar (Jhogar) considera Jhogar = 1 si la observación corresponde a un individuo que se reporta como jefe del hogar y 0 en caso contrario. Se observa que la duración del desempleo es menor en el caso de jefes de hogar. En particular, el 50% de los hombres han conseguido empleo a los 5 meses, en tanto que las mujeres con similar condición gastarían, para lograrlo, 18 meses. Las pruebas estadísticas de Wilcoxon, en cada caso, indican diferencia estadística en la duración.

ii) Variable menores a cargo (menores_a_cargo): Dicha variable indica si existen niños menores de 7 años en el hogar, donde 1 significa que la observación corresponde a un individuo en cuyo hogar existen menores a cargo y 0 en caso contrario. La menor duración del desempleo se reporta en el caso de hombres con menores a cargo. Como se observa en la tabla 1, esta variable no marca diferencia en el caso de las mujeres. Al considerar el total y la submuestra conformada por hombres se tiene diferencia estadística de la duración. Resultado que no se alcanza en el caso de la muestra conformada por mujeres.

Gráfico 3. Análisis de la duración del desempleo por género.



Fuente: elaboración propia

iii) Variable edad: A partir de esta se conforman 5 grupos de individuos. Así se tiene: edad_1 si el individuo es menor de 23 años y 0 en caso contrario, edad_2 si el individuo pertenece al intervalo de edad (22,29] y 0 en caso contrario, edad_3 si el individuo pertenece al intervalo de edad (29, 40] y 0 en caso contrario, edad_4 si el individuo pertenece al intervalo de edad (40,51] y 0 en caso contrario y edad_5 si el individuo es mayor de 51. Así, en general, en los cuantiles 50% y 75%, se observa que a medida que aumenta la edad aumenta la duración del desempleo. Las pruebas estadísticas de diferencia de la duración muestran que el comportamiento entre los grupos es estadísticamente diferente.

iv) Variable mayores a cargo (mayores_a_cargo): Esta variable mide si existen en el hogar estudiantes entre 6 y 22 años de edad, (6,22]. La variable

mayores_a_cargo = 1 indica que la observación corresponde a un individuo en cuyo hogar existen estudiantes con edad en el intervalo referenciado y 0 en caso contrario. Se encuentra una relación directa entre los individuos con personas mayores a cargo y la duración del desempleo. Las pruebas estadísticas de diferencia de la duración, en este caso, sugieren que el comportamiento considerando el total y la submuestra hombres son estadísticamente diferentes.

v) Variable años de educación aprobados a partir de la cual se generan 3 variables *dummy*: aescol1, aescol2 y aescol3 asociadas a menos de 6 años aprobados, entre 6 y 11 años aprobados, respectivamente. En este caso, se observa, para el cuantil 50%, de manera general, que la duración del desempleo aumenta en la medida en que se tiene mayor nivel de escolaridad. Las pruebas estadísticas de diferencia de la

duración muestran que el comportamiento entre los grupos es estadísticamente diferente.

vi) Variable ingreso familiar total real neto³: A partir de esta variable se conforman 8 grupos de individuos por ingreso. Así se generan 8 variables *dummy* desde ingfamtotr_1 hasta ingfamtotr_8 que corresponden a los siguientes intervalos de ingreso:

menor a \$200.000,
 (\$200.000, \$500.000],
 (\$500.000, \$888.827],
 (\$888.827, \$1.593.026],
 (\$1.593.026, \$2.676.212],
 (\$2.676.212, \$3.768.786],
 (\$3.768.786, \$6.992.029]
 y mayores a \$6.992.029.

Es de señalar que, en general, no se observa un patrón de comportamiento claro entre los ingresos y la duración del desempleo. Las pruebas estadísticas de diferencia de la duración permiten concluir que el comportamiento entre los grupos es estadísticamente diferente.

³ Sin considerar el ingreso del individuo cuando este es un individuo ocupado.

vii) Variable estado civil, (est-civil): Se tiene que estcivil = 1 si la observación corresponde a una persona casada (o en unión) y 0 en caso contrario. Se observa, en el caso de la submuestra conformada por hombres, que la duración del desempleo es menor siendo casado; el 50% de los hombres casados salen del estado de desempleo en el quinto mes.

Contrariamente, en el caso de las mujeres, la condición de estar casada aumenta la duración del desempleo y el 50% de ellas requieren de 24 meses para abandonar el estado de desempleo. Las pruebas estadísticas de diferencia de muestras reportan que el comportamiento entre los dos grupos es estadísticamente diferente.

5. Modelación econométrica

El análisis econométrico se centra en dos aspectos fundamentales del desempleo en Colombia. El primero, su duración. Con tal propósito se estima el valor esperado de la duración del desempleo a partir de un modelo de regresión paramétrica de duración, donde el logaritmo de la duración se explica

Tabla 1. Estimación de diferencias en la duración del desempleo entre personas con distintas características.

Características		Total				p-value Test	Hombres				p-value Test	Mujeres			
		Cuantiles			p-value Test		Cuantiles			p-value Test		Cuantiles			p-value Test
		25%	50%	75%			25%	50%	75%			25%	50%	75%	
Género	Mujer	5	24	60	0.0001										
	Hombre	1	10	52											
Jefe de Hogar	No	4	24	60	0.0001	2	12	52	0.0001	6	24	60	0.0001		
	Si	1	9	48		0.5	5	48		3	18	60			
Menores a Cargo	No	3	22	60	0.0001	2	12	60	0.0001	5	24	60	0.9896		
	Si	2	12	48		0.5	5	36		5	24	57			
Edad	<23	3	12	36	0.0010	2	12	36	0.0001	5	18	36	0.0228		
	(22,29]	3	13	42		1	7	36		5	18	48			
	(29,40]	2	12	52		0.5	5	48		5	24	60			
	(40,51]	2	24	60		1	8	52		5	24	64			
	>51	5	36	86		3	36	86		7	36	72			
Mayores a Cargo	No	3	15	58	0.0022	1	9	52	0.0003	5	24	60	0.8159		
	Si	3	18	60		2	11	60		5	24	60			
Grados	<6	1	10	48	0.0001	1	5	36	0.0001	3	18	60	0.0001		
	(5,11]	3	15	48		1	8	48		5	24	58			
Aprobados	>11	6	36	60		4	36	96		3	36	60			
Ingreso Neto (miles)	<200	0.5	5	24	0.0001	0.5	3	12	0.0001	1	8	48	0.0001		
	(200,500]	3	18	60		1	9	60		6	24	60			
	(500,888]	4	24	60		2	12	52		7	24	60			
	(888,1593]	4	24	60		2	12	60		6	24	60			
	(1593,2676]	4	24	60		2	18	96		6	24	60			
	(2676,3768]	4	24	72		3	24	84		7	36	60			
	(3768,6992]	5	36	84		6	48	.		5	24	60			
>6992	6	48	.	8	.	.	4	48	.						
Estado Civil	No Casado	3	18	60	0.0001	3	15	60	0.0001	4	22	60	0.0001		
	Casado	2	14	52		0.5	5	48		6	24	60			

Fuente: elaboración propia

mediante un conjunto de variables comúnmente utilizadas en la literatura; y el segundo, enmarcado en su probabilidad. Así, con base en modelación Probit y un conjunto de variables explicativas se estudia la probabilidad de que un individuo esté en situación de desempleo.

5.1. Modelo de duración del desempleo

5.1.1. Conceptos básicos y conformación de la muestra.

Se define la variable aleatoria T como la duración del desempleo, la cual mide el tiempo transcurrido en semanas desde que el individuo abandonó su último empleo y entró al estado de desempleo. La estimación de dicha duración se lleva a cabo mediante un modelo de regresión paramétrica. En particular, se utiliza el modelo *Accelerated Failure Time* o modelo AFT.

El modelo de regresión AFT es especificado a partir de la ecuación (10)

$$\ln(t) = X' \beta + u \quad (10)$$

De tal forma que diferentes distribuciones del término de perturbación, como lo son las distribuciones weibull o lognormal, conducen a diferentes modelos AFT. Así, la duración puede expresarse a partir de la ecuación (11)

$$t = \exp(X' \beta) v \quad (11)$$

$$v = e^u \quad (12)$$

Considerando este modelo, existen diferentes formas de llevar a cabo el pronóstico de la duración del evento bajo consideración (Cleves, 2008). En particular, el pronóstico se puede alcanzar a través de la siguiente ecuación:

$$\hat{t}_i = \exp(X_i' \beta + E[u_i])$$

Donde $E[u_i] = \frac{\Gamma(1/\rho)}{\rho}$ siendo ρ el parámetro de escala de la distribución Weibull y $\Gamma(1)$ la distribución Gamma evaluada en el valor 1.

5.2. Modelo de probabilidad

El modelo Probit es un caso particular de los modelos de selección discreta donde la variable dependiente y_i es binaria; solo toma dos valores 0 si el individuo está empleado y 1 si este evento no ocurre (Cameron y Trivedi, 2005; Baltagi, 2001; Wooldridge, 2001, y Maddala, 1997). Existe un proceso generador de respuesta y_i^* el cual es no-observado en la realidad y está definido como la relación siguiente:

$$y_i^* = x_i \beta + u_i \quad (13)$$

$$E[y_i^*] = x_i \beta$$

Lo que se observa en cambio es la variable binaria y_i :

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (14)$$

donde $E[y_i] \neq x_i \beta$.

El objetivo es determinar la probabilidad de ocurrencia de un evento condicionado a las características de los individuos de la muestra. La probabilidad de que el evento ocurra está definida como

$$P_i = \text{Prob}(y_i = 1) \quad (15)$$

$$\text{Prob}(y_i = 1) = \text{Prob}(u_i > -x_i \beta)$$

$$= 1 - F_i(-x_i \beta)$$

$$P_i = F_i(x_i \beta)$$

$$x_i = \{\text{características del individuo } i\}$$

$$\beta \rightarrow \text{«Ponderadores por cada característica»}$$

El modelo probit surge de suponer que $u_i \sim N(0, \sigma^2)$. En particular, en este trabajo se revisa la bondad de ajuste del modelo a los datos observados a través de la técnica de calidad de la predicción. Así, se construye una variable indicadora \hat{y}_i

$$\hat{y}_i = \begin{cases} 1 & \text{si } P_i > h \\ 0 & \text{si } P_i \leq h \end{cases} \quad (16)$$

Donde $0 < h < 1$ es el umbral tal que si la probabilidad estimada supera el umbral la variable indicadora \hat{y}_i toma el valor de 1. Se construye otra variable indicadora w_i donde toma el valor de 1 si el modelo pronosticó correctamente y, de lo contrario, toma el valor de 0:

$$w_i = \begin{cases} 1 & \text{si } y_i = \hat{y}_i \\ 0 & \text{si } y_i \neq \hat{y}_i \end{cases} \quad (17)$$

La calidad de predicción (C_p) dado un umbral h es el promedio del número de veces que el modelo pronostica correctamente.

$$C_p|h = \frac{1}{N} \sum_{i=1}^N w_i \quad (18)$$

6. Conjunto de información relevante en la estimación econométrica

6.1. Descripción de variables

El modelo de duración del desempleo se estima a partir del modelo AFT, explicado anteriormente, donde la variable dependiente hace referencia a la duración del desempleo como se define a continuación:

6.1.1. Duración del desempleo.

Hace referencia a la dupla de variables que informan sobre cada individuo la duración de su desempleo y su categorización como individuo censurado o no censurado. Es decir, si el individuo ya completó la duración de su desempleo o no, las variables que conforman este grupo son:

- Duración: La variable indica cuánto tiempo lleva desempleado en caso de no haber conseguido empleo o cuánto tiempo se demoró en conseguir el em-

pleo si el individuo ya cambió de estado (es decir, consiguió un empleo).

- **Censura:** Esta variable indica si la observación está en censura o ya culminó su período de desempleo

En particular, en este trabajo, el conjunto de individuos que culminaron su período de desempleo se conforma con la posición del mes dentro del trimestre, como se explicó anteriormente.

El conjunto de variables explicativas o covariables que conforman la matriz de diseño X , para cada una de las muestras mencionadas, se definen a continuación:

6.1.2. Edad

La duración del desempleo puede ser explicada, en parte, por la edad de cada individuo. En principio esta variable es un indicador de la experiencia y del conocimiento que el individuo tenga del mercado laboral, pero también está correlacionada con otra serie de factores que pueden afectar el comportamiento del individuo y los resultados de sus acciones. Muchos de estos aspectos no están linealmente relacionados con la probabilidad de desempleo o con la duración de la búsqueda de empleo. Por esa razón, la edad se incluye en las estimaciones como una forma cuadrática (edad y edad al cuadrado). Esto tiene dos efectos: primero, posiblemente los efectos marginales van a depender de los niveles de dicha variable y, segundo, los test de significancia usuales no son aplicables. En vez de ellos se deben usar test conjuntos como la razón de verosimilitud.

6.1.3. Educación

La variable años aprobados de educación (*aeduc*) permite capturar el efecto de la especialización sobre la duración del desempleo.

Esta variable es de carácter continuo; al igual que la edad, se introduce en forma cuadrática a fin de capturar efectos más complejos.

6.1.4. Reemplazo.

El grupo de variables denominadas de reemplazo son todas aquellas variables que permitan explicar el salario de reserva de un individuo. Así, se incorpora de manera indirecta el efecto del salario de reserva sobre la duración del desempleo. Se denominan variables de reemplazo porque son todos aquellos factores que pueden determinar la urgencia que tiene un individuo de conseguir empleo.

- **Estado civil (*estcivil*):** Es muy posible que una persona casada tenga mayores responsabilidades implicando un salario de reserva más bajo. La variable toma valor de 1 si el individuo está casado o vive en unión libre.
- **Jefe de hogar (*jhogar*):** Al igual que la variable anterior, ser jefe de hogar puede estar relacionado negativamente con el salario de reserva.
- **Número de menores a cargo (*FNmenor*):** Variable de *responsabilidad económica* sobre otros miembros de la familia⁴ que afecta el salario de reserva en dos vías contrarias. Puede suceder que, entre más personas a cargo se tengan, mayor es el salario de reserva porque su costo de oportunidad es mayor o puede ser que entre más personas a cargo se tenga, su salario de reserva es menor dado que se tiene más necesidad de trabajar.
- **Ingreso del resto de la familia, *ing_res_fam*,** resta del *Ingreso_familiar_total* el ingreso laboral del individuo ocupado. Podría entenderse como el ingreso del resto de la familia y determina el salario de reserva de cada individuo. Posiblemente es un *proxy*

del costo de los recursos disponibles para financiar la búsqueda de empleo y en ese sentido está asociado con el salario de reserva del individuo. Es de señalar que dicha variable se considera bajo la transformación logarítmica, *ling_res_fam*.

- **Otros controles:** Género (hombres = 1, mujeres = 0).

6.2. Construcción de las variables asociadas a las condiciones del mercado referentes a duración del desempleo y salarios por ocupación

Adicionalmente, se consideran las condiciones de mercado en lo que respecta a duración del desempleo (*dur_cm*) y al salario por ocupación (*ing_cm*), se considera en el modelo su transformación logarítmica (*ling_cm*). La construcción de estas dos variables se lleva a cabo siguiendo la metodología explicada a continuación:

En la literatura internacional se encuentra la sugerencia de utilizar como variable explicativa del modelo de duración del desempleo a las expectativas de dicha duración. Esta variable es de gran importancia porque permite modelar el umbral en el cual las personas empiezan a ajustar sus estándares. Así, si un individuo ha superado el tiempo esperado de desempleo entonces aceptará ofertas laborales con mayor facilidad, buscará otras opciones u otros cargos para disminuir el tiempo de búsqueda de empleo. En este artículo se usará como variable *proxy* de las expectativas de duración a las condiciones del mercado respecto a dicha variable.

La construcción de la variable *proxy* de expectativas se lleva a cabo a través de métodos paramétricos que permiten aproximar una distribución conocida a los datos observados de la duración del desem-

⁴Véase anexo 3 (glosario de variables).

pleo. Así, con base en los cuantiles de la distribución se llega a inferir las condiciones del mercado sobre el tiempo de desempleo para cada una de las ramas de actividad reportadas en la encuesta⁵.

En esta forma, se considera la información correspondiente a duración del desempleo y la ocupación para la muestra de ocupados, con menos de tres meses de ocupación respecto a cada uno de los meses que conforman el tercer trimestre de 2010: Posteriormente, se aproxima para cada una de estas ocupaciones una distribución asociada al tiempo reportado de desempleo. Es de señalar que se tiene información a priori acerca de la distribución que pueden seguir los datos; esta debe ser positiva y continua, dada la naturaleza del problema. La estadística sugiere (Evans et al., 2011) como distribuciones posibles las distribuciones gamma, log-normal, Chi-cuadrado, exponencial, F (snedecor) y Weibull. Los parámetros se estiman por el método de máxima verosimilitud (ecuación 37), y una vez se determina la distribución que siguen los datos se construyen las expectativas de la duración del desempleo en los cuantiles de interés.

$$\text{Max}_{\mu, \sigma} \prod_{i=1}^T f(X_i | \mu, \sigma) \quad (19)$$

La selección de la distribución se lleva a cabo a través de una medida de evaluación del ajuste de esta a los datos. Es decir, se selecciona aquella que presente la menor diferencia entre los datos observados y la distribución estimada (d_i). En particular, como criterio de selección, se utiliza el método de error cuadrático medio (ECM)⁶ (ecuación 20).

$$\text{ECM} = \frac{\sum_{i=1}^T d_i^2}{T} \quad (20)$$

Aplicando esta metodología a la duración del desempleo por ocupación se encuentra que las distribuciones que mejor se aproximan al tiempo de la duración del desempleo por ocupación son las distribuciones Gamma y Weibull⁷. Para la construcción de la variable de condiciones del mercado referentes al tiempo de duración del desempleo por ocupación se toma el cuantil 75, dado que este exhibe una mayor concentración de masa de probabilidad. Así, este es el umbral de duración donde el 75% de los individuos de la muestra consiguen empleo.

Se encuentra la Gamma como la mejor distribución que aproxima al salario mensual por ocupación. Para la construcción de la variable de condiciones del mercado en lo referente al salario por ocupación se toma el cuantil 75 dado que este exhibe una mayor concentración de masa de probabilidad. Un individuo con expectativas elevadas de salario puede ser que se demore más tiempo en conseguir empleo y considere, en algún momento, cambiar de oficio para alcanzar el nivel deseado. Para cada una de las ramas de ocupación de la muestra se exhiben expectativas de salario diferentes que están relacionadas con el perfil necesario en cada ocupación.

Es de resaltar que los resultados de las condiciones del mercado en lo referente al cuantil 75, tanto de la duración del desempleo (meses) como del salario (salario medido en pesos deflactado por el índice de precios de cada mes), se construyen por cada ocupación, teniendo

en cuenta género, edad⁸ y años de educación⁹.

7. Resultados de los modelos econométricos

Las estimaciones econométricas se llevan a cabo con base en la información contenida en la muestra total y en cada una de las submuestras conformadas por hombres y mujeres, respectivamente. En particular, se conforman cuatro matrices de diseño y, por consiguiente, cuatro modelos por estimar:

- $X_1 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2}\}$
- $X_2 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, jhogar, estcivil, FNmenor}\}$
- $X_3 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, ling_cm, dur_cm}\}$
- $X_4 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, jhogar, estcivil, FNmenor, ling_cm, dur_cm}\}$

7.1. Resultados de la estimación de los modelos de duración

La estimación de los diferentes modelos de duración se lleva a cabo a través del modelo AFT, explicado en la sección 6.2. En las distintas estimaciones se considera que el término de perturbación sigue una distribución Weibull. Es de señalar que la decisión acerca de la distribución se toma con base en pruebas realizadas considerando otras distribuciones¹⁰. En particular, se considera la distribución exponencial como un caso particular de la Weibull, y se rechaza su existencia mediante un test de multiplicadores de Lagrange.

⁵ Surge la pregunta: ¿por qué no utilizar el tiempo promedio del desempleo para cada uno de los oficios? Y la respuesta es que debido a la alta dispersión presentada, el promedio no daría una medida de tendencia central confiable.

⁶ Véase un ejemplo particular en el anexo 2.

⁷ Es de señalar que la distribución Weibull es un caso particular de la distribución Gamma ante modificaciones en el parámetro de forma.

⁸ Grupo 1: menores a 30 años; grupo 2: entre 30 y 50 años, y grupo 3: mayores a 50 años.

⁹ Grupo 1: menos de 6 años de educación; grupo 2: entre 6 y 11 años de educación, y grupo 3: más de 11 años de educación.

¹⁰ Entre otras, Lognormal y Exponencial.

La estimación de los cuatro modelos se realiza sobre la muestra total y sobre las submuestras construidas considerando la información asociada a hombres y a mujeres.

En particular, al considerar la información total, los diferentes modelos incluyen la variable género. En el caso de las submuestras dicha variable es excluida.

Las tablas 2, 3 y 4 presentan los resultados de la estimación de los cuatro modelos considerando la muestra total y las submuestras generadas por género, respectivamente.

Tabla 2. Estimación Probit muestra total.

Muestra Total								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value
Intercept	2.2324	<.0001	2.0611	<.0001	3.2392	<.0001	3.0957	<.0001
genero	-0.7499	<.0001	-0.7016	<.0001	-0.479	<.0001	-0.4466	<.0001
edad	-0.068	<.0001	-0.0402	<.0001	-0.0487	<.0001	-0.024	0.002
edad2	0.0012	<.0001	0.0009	<.0001	0.0008	<.0001	0.0006	<.0001
aeduc	0.0815	<.0001	0.0753	<.0001	0.0516	0.0009	0.0486	0.0016
aeduc2	0.0003	0.7367	0.0001	0.8916	0.0015	0.0869	0.0012	0.161
ling_res_fam	0.0673	0.0006	0.0983	<.0001	0.0678	0.0004	0.0959	<.0001
ling_res_fam2	0.0017	0.2023	-0.0011	0.418	0.0015	0.2602	-0.0011	0.4203
ling_cm					-0.1198	<.0001	-0.1188	<.0001
dur_cm					0.0445	<.0001	0.0414	<.0001
jhogar			-0.5827	<.0001			-0.5467	<.0001
estcivil			-0.1647	<.0001			-0.1441	<.0001
FNmenor			-0.0903	<.0001			-0.0901	<.0001
Scale	1.4444		1.4311		1.4148		1.4037	
Weibull Shape	0.6923		0.6988		0.7068		0.7124	

Fuente: elaboración propia

Las estimaciones de la tabla 2 arrojan resultados interesantes. En general se encuentra que los hombres tienen procesos de búsqueda de empleo más cortos que las mujeres. Esto se puede explicar porque las mujeres, al ser las principales responsables por las tareas domésticas y el cuidado de los niños, tienen mayores demandas sobre su tiempo y, por lo tanto, un mayor salario de reserva que los hombres (otras cosas iguales), según lo indica la teoría presentada anteriormente.

También se encuentra que las personas casadas y los jefes de hogar tienen desempleo de más corta duración que el resto de la población.

Esto es consistente con la hipótesis de que dichas personas, por ser los principales responsables del sostenimiento del hogar, tienen salarios de reserva menores y, por lo tanto, menores períodos de búsqueda de empleo (son menos exigentes con respecto a las condiciones de trabajo).

La combinación de signos de la variable edad (negativo para el primer término y positivo para el término cuadrático) indica que la duración del desempleo disminuye con la edad hasta cierto punto, pero a partir de dicho punto aumenta. La edad crítica (punto donde la duración del desempleo deja de decrecer y comienza a aumentar) está entre 20 y 30 años de edad, dependiendo del modelo que se esté usando¹¹. Esto

en general implica que, otras cosas iguales, la duración del desempleo de los jóvenes es menor que la de las personas mayores.

Los años de educación (aeduc) tienen un efecto cuadrático que aumenta la duración del desempleo. Las personas con bajos niveles de educación tienen períodos de desempleo de baja duración; pero a medida que los niveles educativos aumentan, los procesos de búsqueda se hacen más prolongados. Esto se puede deber a que entre mayores sean los niveles educativos, mayor la complejidad de los empleos y, por lo tanto, mayor la información que los buscadores necesitan obtener antes de aceptar un empleo. Esto hace que los procesos de búsqueda sean más prolongados.

¹¹ Este punto está dado por los coeficientes de edad y edad² = - edad / (2*edad²).

Tabla 3. Estimación Probit submuestra hombres.

Sub muestra: Hombres								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value
Intercept	2.3275	<.0001	1.7465	<.0001	2.4618	<.0001	1.8416	<.0001
edad	-0.1228	<.0001	-0.0589	<.0001	-0.0917	<.0001	-0.0346	0.0009
edad2	0.0019	<.0001	0.0013	<.0001	0.0013	<.0001	0.0008	<.0001
aeduc	0.0737	0.0008	0.0687	0.0012	0.051	0.0162	0.0497	0.0158
aeduc2	0.0027	0.0325	0.0019	0.1099	0.002	0.0895	0.0013	0.2775
ling_res_fam	-0.0103	0.6978	0.0318	0.22	0.003	0.9082	0.0426	0.0906
ling_res_fam2	0.0069	0.0002	0.0031	0.0916	0.0056	0.0018	0.002	0.2541
ling_cm					-0.0572	0.0007	-0.0478	0.0034
dur_cm					0.0925	<.0001	0.0863	<.0001
jhogar			-0.4059	<.0001			-0.3771	<.0001
estcivil			-0.6686	<.0001			-0.6163	<.0001
FNmenor			-0.1362	<.0001			-0.1337	<.0001
Scale	1.468		1.4335		1.4183		1.3882	
Weibull Shape	0.6812		0.6976		0.7051		0.7203	

Fuente: elaboración propia

Tabla 4. Estimación Probit submuestra mujeres.

Sub muestra: Mujeres								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value	Coefficiente	P-value
Intercept	1.2054	<.0001	1.3754	<.0001	2.7168	<.0001	2.851	<.0001
edad	0.0035	0.7538	-0.0069	0.5532	0.034	0.0023	0.023	0.0506
edad2	0.0002	0.16	0.0004	0.0104	-0.0003	0.08	-0.0001	0.6585
aeduc	0.0775	0.0005	0.0769	0.0005	0.0292	0.1998	0.0301	0.1845
aeduc2	-0.0015	0.1975	-0.0014	0.2246	0.0015	0.2219	0.0015	0.2213
ling_res_fam	0.1548	<.0001	0.1772	<.0001	0.1515	<.0001	0.1709	<.0001
ling_res_fam2	-0.0044	0.0237	-0.0065	0.0009	-0.0043	0.0262	-0.0061	0.0016
ling_cm					-0.1644	<.0001	-0.1601	<.0001
dur_cm					0.0224	<.0001	0.0214	<.0001
jhogar			-0.3566	<.0001			-0.3256	<.0001
estcivil			0.2333	<.0001			0.2196	<.0001
FNmenor			-0.007	0.7942			-0.0117	0.6587
Scale	1.3951		1.3865		1.3772		1.3705	
Weibull Shape	0.7168		0.7212		0.7261		0.7297	

Fuente: elaboración propia

El ingreso del resto de la familia (ling_res_fam), como ya se dijo, está relacionado con la disponibilidad de recursos para financiar la búsqueda de empleo¹². En ese sentido, entre mayor sea el ingreso del resto de la familia, mayor será el salario de reserva de la persona y (otras cosas iguales) mayor será la duración de la búsqueda de empleo. Esta relación se conoce como el «efecto aspiraciones». Sin embargo, en un país como Colombia en donde la mayor parte de los trabajadores consiguen empleo a través de contactos, amigos, conocidos, etc., el ingreso del resto de la familia también es un proxy de la calidad de dichas redes. El efecto de la calidad de los contactos es hacer más eficientes los procesos de búsqueda y, por consiguiente, mejorar las ofertas laborales que el trabajador hace, reduciendo de esta forma la duración esperada del desempleo. Este segundo efecto lo podemos llamar «efecto oportunidades».

Los resultados de la tabla 2 indican que en general predomina el efecto aspiraciones, aunque los signos negativos de la forma cuadrática del modelo 2 y el modelo 4 son evidencia de que a niveles altos de ingreso el efecto oportunidades puede ser importante.

Es importante mencionar los resultados de las variables que miden las condiciones del mercado laboral como son el salario esperado (Ling_cm) y la duración esperada de la búsqueda de empleo (dur_cm).

Los resultados de la tabla 2, modelos 3 y 4, indican que entre mayor sea el salario esperado (medido por lo que reciben los trabajado-

res que han conseguido empleo recientemente) menor será la duración del desempleo. Esto se puede interpretar en el sentido de que las ocupaciones que (otras cosas iguales) pagan mejores salarios probablemente son ocupaciones donde hay presiones de demanda y por lo mismo la consecución de empleo en ellas es más fácil. Posiblemente los empleadores hacen mayores inversiones en la generación de información para facilitar el enganche de trabajadores.

La duración esperada del desempleo (dur_cm) arroja signo positivo, lo cual indica que en las ocupaciones donde es más difícil conseguir empleo, los trabajadores que buscan dicho empleo están dispuestos a hacer búsquedas más prolongadas.

En conjunto, los resultados de estas dos últimas variables muestran que las condiciones de los mercados (por ocupaciones) son importantes en los procesos de búsqueda de empleo.

Los resultados de las tablas 3 y 4 para hombres y mujeres, respectivamente, confirman en general los

resultados anteriores, con algunas excepciones que mencionamos a continuación:

Tal vez el principal cambio en los resultados cuando se desagrega por sexo es el efecto del estado civil. Para los hombres esta variable sigue teniendo un efecto negativo, pero para las mujeres es positivo. Este resultado es consistente con la hipótesis ya planteada de que (otras cosas iguales) las mujeres tienen salarios de reserva más altos que los hombres y por eso sus procesos de búsqueda son más exigentes y prolongados. El hecho de que el estar casadas tenga un efecto positivo sobre la duración del desempleo de las mujeres indica precisamente esto.

7.2. Resultados de la estimación de los modelos de probabilidad

A continuación se presentan los resultados de la estimación del modelo Probit robusto, el cual está corregido por la heterogeneidad no observada.

La muestra en los modelos de probabilidad está conformada por el total de ocupados y desocupados

Tabla 5. Estimación Probit robusto muestra total.

Muestra: Total				
	Modelo 1		Modelo 2	
	Coefficiente	P-value	Coefficiente	P-value
Intercept	0.1165	0.0120	-0.0652	0.1820
genero	-0.2221	0.0000	-0.1633	0.0000
edad	-0.0633	0.0000	-0.0467	0.0000
edad2	0.0006	0.0000	0.0004	0.0000
aeduc	0.0176	0.0000	0.0179	0.0000
aeduc2	-0.0009	0.0000	-0.0009	0.0000
ling_res_fam	0.0395	0.0000	0.0512	0.0000
ling_res_fam	-0.0013	0.0030	-0.0025	0.0000
jhogar			-0.2326	0.0000
estcivil			-0.1683	0.0000
FNmenor			0.0016	0.8080

Fuente: elaboración propia

¹² En ausencia de seguros de desempleo, los trabajadores dependen de sus propios ingresos y ahorros para financiar la búsqueda de empleo. En general, los mercados financieros no generan mecanismos para financiar los procesos de búsqueda de trabajo (fallas de mercado).

¹³ El umbral se define como un valor en el intervalo (0,1) que mide la proporción de aciertos del pronóstico versus lo observado.

Tabla 6. Estimación Probit robusto submuestra hombres.

Sub muestra: Hombres				
	Modelo 1		Modelo 2	
	Coefficiente	P-value	Coefficiente	P-value
Intercept	-0.1325	0.0340	-0.4476	0.0000
edad	-0.0704	0.0000	-0.0398	0.0000
edad2	0.0007	0.0000	0.0005	0.0000
aeduc	0.0110	0.1000	0.0119	0.0780
aeduc2	-0.0002	0.6260	-0.0002	0.4710
ling_res_fan	0.0165	0.0570	0.0483	0.0000
ling_res_fan	0.0008	0.2060	-0.0020	0.0010
jhogar			-0.3380	0.0000
estcivil			-0.2581	0.0000
FNmenor			-0.0453	0.0000

Fuente: elaboración propia

Tabla 7. Estimación Probit robusto submuestra mujeres.

Sub muestra: Mujeres				
	Modelo 1		Modelo 2	
	Coefficiente	P-value	Coefficiente	P-value
Intercept	-0.0313	0.6570	-0.1010	0.1640
edad	-0.0440	0.0000	-0.0394	0.0000
edad2	0.0002	0.0000	0.0002	0.0000
aeduc	0.0219	0.0020	0.0228	0.0010
aeduc2	-0.0015	0.0000	-0.0015	0.0000
ling_res_fan	0.0640	0.0000	0.0646	0.0000
ling_res_fan	-0.0034	0.0000	-0.0035	0.0000
jhogar			-0.0718	0.0000
estcivil			-0.0617	0.0000
FNmenor			0.0339	0.0000

Fuente: elaboración propia

de la Gran Encuesta Integrada de Hogares, agregación de los meses correspondientes al tercer trimestre de 2010. La variable dependiente es 1 si la persona está desempleada y cero si está ocupada. Es de señalar que tan solo se consideran las matrices de diseño X_1 y X_2 . Es decir, no se consideran las condiciones de mercado en lo referente a duración y salario.

En las tablas 5, 6 y 7 se consignan los resultados de la estimación del modelo Probit robusto para la muestra total y las submuestras construidas mediante la variable género, respectivamente. En la primera columna de cada una de las muestras y de cada uno de los modelos se encuentran los coeficientes estimados, los cuales, dada la no linealidad, no son directamente interpretables. La segunda columna es el *P-value* asociado a los coeficientes estimados.

¹⁴ Se tienen tres submuestras de datos definidas a partir del tiempo que llevan desempleados, según el mes del tercer trimestre.

El gráfico 4 muestra la calidad de predicción de los modelos 1 y 2 para el total de la muestra. En general, el modelo 1 pronostica correctamente el 67% de las veces, en tanto que el segundo modelo lo hace el 68% de las veces. La línea azul representa la calidad de la predicción para cada uno de los umbrales en el eje horizontal¹³. La línea delgada es la calidad de pronóstico de una variable aleatoria con distribución Bernoulli.

El gráfico 5 muestra la calidad de predicción de los modelos 1 y 2 para la sub muestra conformada por hombres. El modelo 1 pronostica correctamente el 66% de las observaciones, en tanto que el segundo modelo lo hace para el 69,4% de estas.

El gráfico 6 muestra la calidad de predicción de los modelos 1 y 2 para la submuestra conformada por mujeres. Los dos modelos pronostican correctamente el 67% de las observaciones.

7.2.1. Perfiles.

a. Duración. A continuación se presentan los pronósticos de la duración del desempleo por perfiles de individuo, considerando los parámetros estimados en cada una de las submuestras (hombres y mujeres). Es de señalar que se consideran los promedios de las variables explicativas de cada modelo de duración correspondiente a cada submuestra particular. Así, el promedio del ingreso familiar total y la variable que mide menores a cargo son obtenidos sobre cada una de las submuestras¹⁴ seleccionando por estado civil, jefe de hogar y edad.

La tabla 8 presenta los resultados de los pronósticos de la duración del desempleo por perfil considerando la submuestra conformada por hombres. Los cuatro modelos considerados.

Gráfico 4. Muestra total.

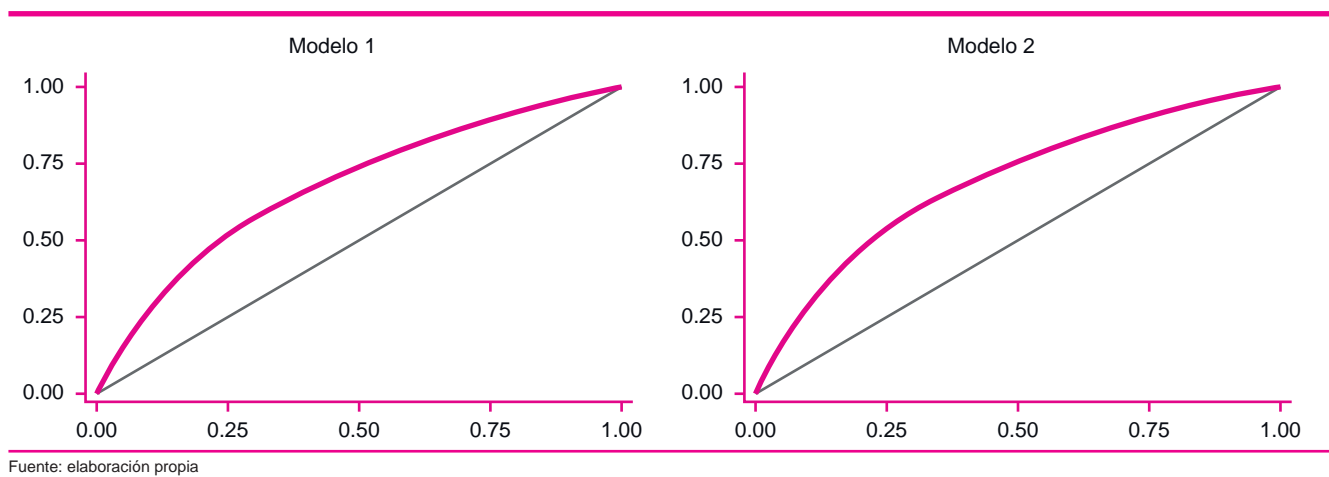


Gráfico 5. Submuestra hombres.

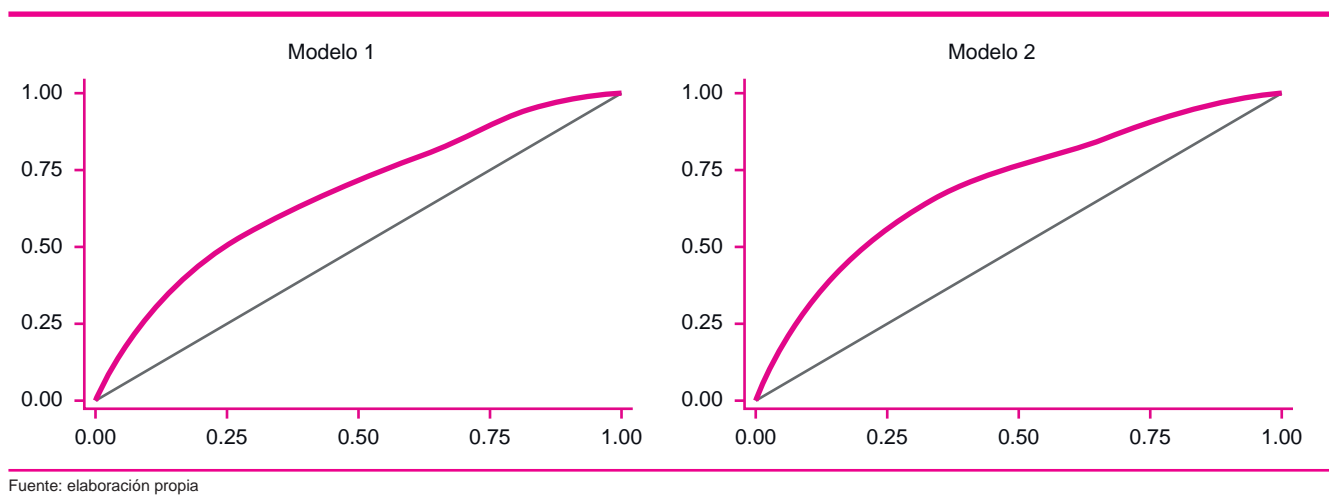
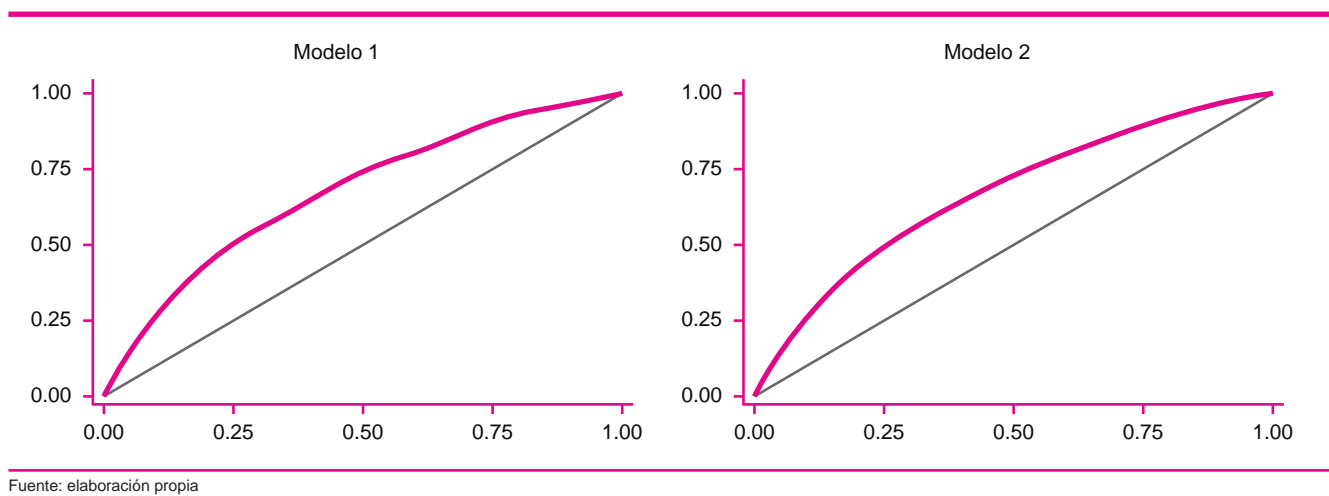


Gráfico 6. Submuestra mujeres.



En la tabla 9 se consignan los resultados de la duración del desempleo por perfil considerando la muestra de mujeres.

Como se observa, de manera general, la duración del desempleo es creciente con la edad y la edu-

cación. Así, una persona que pierde el empleo tarda más tiempo en conseguir un nuevo empleo entre más edad y mayor educación presente. Igualmente, se concluye que las medidas de duración correspondientes a las mujeres son siempre mayores a las reportadas

por los hombres. Es decir, se alcanzan resultados robustos frente a la selección de la muestra y a los diferentes modelos. Es de señalar que el modelo 4 reporta las menores duraciones del desempleo en las dos submuestras consideradas.

Tabla 8. Hombres.

Jefe de Hogar=0, Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	2.2	3.3	7.1	16.2	2.0	3.0	6.4	15.4	2.6	4.7	8.7	19.5	4.9	8.2	16.3	36.0
Modelo 2	3.3	4.9	9.5	19.7	4.1	6.2	11.5	23.0	6.4	10.8	18.5	36.0	12.6	19.8	36.0	70.0
Modelo 3	2.3	3.3	5.5	10.4	2.1	2.9	4.7	8.8	2.4	3.5	5.7	9.7	3.7	5.3	9.3	15.1
Modelo 4	3.3	4.8	7.5	13.2	4.1	5.7	8.4	13.8	5.8	8.1	12.2	19.0	9.4	12.9	21.2	31.7
Jefe de Hogar=0, Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	1.9	3.1	6.8	15.6	1.8	2.9	6.3	15.4	2.5	4.1	8.7	19.3	5.0	8.2	17.4	44.4
Modelo 2	1.4	2.3	4.4	9.3	1.9	2.9	5.3	11.1	3.2	4.9	9.1	17.8	6.5	10.2	19.1	41.6
Modelos	2.2	3.0	5.2	9.8	1.7	2.7	4.5	9.0	2.2	3.2	5.7	9.6	3.7	5.2	9.7	18.3
Modelo 4	1.6	2.3	3.6	6.4	1.9	2.7	4.0	7.1	2.8	4.0	6.4	9.8	5.0	6.9	11.6	19.6
Jefe de Hogar=1, Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	0.6	3.1	5.9	13.2	1.8	2.7	5.3	11.9	2.5	4.0	7.5	16.3	4.9	6.9	14.9	29.0
Modelo 2	0.9	3.4	5.8	11.8	2.9	4.0	7.0	13.4	4.3	6.6	11.4	21.7	8.7	12.3	23.3	40.3
Modelo 3	0.8	3.0	4.5	8.1	1.7	2.4	3.9	6.8	2.1	3.1	4.9	8.7	4.0	4.8	8.6	12.0
Modelo 4	1.1	3.4	4.7	7.9	2.7	3.6	5.4	8.2	3.7	5.2	7.7	12.6	7.2	8.7	14.3	18.5
Jefe de Hogar=1, Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	1.6	2.8	5.8	12.9	1.7	2.7	5.6	14.4	2.4	3.8	8.0	19.3	5.0	8.0	15.9	46.2
Modelo 2	0.9	1.4	2.7	5.5	1.2	1.9	3.5	7.3	2.0	3.1	5.9	11.8	4.4	6.7	12.1	28.1
Modelos	11.9	2.8	4.4	8.1	1.7	2.5	4.1	8.1	2.1	3.0	5.2	10.0	3.6	5.0	8.9	19.6
Modelo 4	1.1	1.5	2.3	4.0	1.3	1.8	2.7	4.6	1.9	2.6	4.2	7.1	3.4	4.6	7.6	14.0

Fuente: elaboración propia

b. Probabilidad del desempleo. Las tablas 10 y 11 presentan los pronósticos de la probabilidad de estar desempleado dado un conjunto de características o perfiles del individuo de

interés: edad, estado civil, años de educación y jefe de hogar. Es de señalar que las restantes variables explicativas se consideran en el promedio de los individuos que conforman la ca-

racterización particular. En este caso, las probabilidades pronosticadas se basan en los modelos 1 y 2, reportados en las tablas 6 y 7. La tabla 10 reporta las probabilidades obtenidas en el caso de

hombres, en tanto que la tabla 11 presenta los resultados para mujeres. Como se destaca, la probabilidad de estar en desempleo es menor entre mayor sea la persona. Dentro de cada edad, en cada uno de los perfiles, los años de educación aumentan la probabilidad de estar en desempleo. En general, la probabilidad de estar desempleado es mayor en el caso de las mujeres.

De las tablas anteriores se puede concluir que, si bien es cierto, una

mayor edad está asociada a una mayor duración del desempleo también está asociada a una menor probabilidad de perder el empleo. Con el fin de concluir al respecto, es necesario construir medidas de incidencia, las cuales se presentan en la siguiente sección.

8. Incidencia del desempleo

Una relación muy importante es la que define la tasa de desempleo

como el producto de la incidencia por la duración del desempleo:

$$\text{Tasa de desempleo mensual} = \text{duración de desempleo (en meses)} \times \text{Tasa de incidencia mensual.}$$

La tasa de incidencia se puede definir como la proporción de la fuerza de trabajo que entra mensualmente a la situación de desempleo.

Esta relación permite distinguir entre diferentes tipos de desempleo: por ejemplo, el desempleo puede

Tabla 9. Mujeres.

Jefe de Hogar=0, Estado Civil=0																
Edad	25				35				45				55			
	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.5	13.3	16.5	7.8	11.2	15.5	18.9	9.9	13.7	18.8	22.9	12.2	17.1	23.4	28.8
Modelo 2	6.3	9.1	12.6	15.4	7.5	10.9	14.9	17.9	10.3	14.0	19.1	23.2	13.8	19.3	25.8	31.3
Modelo 3	6.3	7.9	10.9	12.4	8.0	9.7	12.9	12.9	9.2	11.0	14.5	16.1	10.5	12.9	17.0	18.3
Modelo 4	6.0	7.6	10.5	11.8	7.8	9.4	12.5	12.4	9.5	11.2	14.7	16.4	11.6	14.4	18.6	19.8
Jefe de Hogar=0, Estado Civil=1																
Edad	25				35				45				55			
	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.7	13.5	16.8	8.0	9.9	8.0	9.7	9.6	13.7	18.9	23.2	12.1	17.2	23.6	28.7
Modelo 2	8.1	11.6	16.1	19.8	11.3	13.9	10.1	12.2	12.4	17.5	24.0	29.1	17.3	24.1	32.2	38.9
Modelos	6.6	8.7	11.2	12.4	15.6	18.9	12.9	15.6	9.3	11.4	15.1	15.8	11.5	13.2	17.5	16.5
Modelo 4	8.0	10.3	13.4	14.7	19.1	22.8	13.7	16.4	11.8	14.4	19.0	19.8	15.8	18.0	23.4	22.3
Jefe de Hogar=1, Estado Civil=0																
Edad	25				35				45				55			
	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.3	9.6	13.1	16.1	7.9	11.1	15.3	18.3	9.6	13.5	18.5	22.4	12.1	17.1	23.4	28.3
Modelo 2	4.4	6.5	8.9	11.1	5.5	7.7	10.6	12.8	7.0	9.8	13.5	16.5	9.8	13.6	18.3	22.3
Modelo 3	6.9	8.6	10.7	11.7	7.7	9.4	12.4	13.2	8.9	10.9	14.7	16.0	10.5	12.9	17.7	17.9
Modelo 4	5.0	6.0	7.5	8.4	5.5	6.8	8.9	9.5	6.7	8.2	11.0	12.2	8.6	10.4	14.2	14.4
Jefe de Hogar=1, Estado Civil=1																
Edad	25				35				45				55			
	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.6	13.3	16.3	8.2	11.3	15.4	18.8	9.7	13.6	18.7	22.6	12.1	17.0	23.3	29.2
Modelo 2	5.7	8.2	11.3	13.9	7.1	9.7	13.3	16.2	8.7	12.3	17.0	20.3	12.5	16.7	22.7	27.0
Modelos	7.1	8.0	10.8	12.5	7.8	9.5	13.1	13.0	9.2	10.7	14.6	15.8	11.8	12.4	18.1	14.4
Modelo 4	6.1	7.0	9.4	11.0	6.9	8.4	11.5	11.5	8.4	9.9	13.5	14.5	12.0	12.3	17.8	13.7

Fuente: autor

ser alto porque los procesos de búsqueda de empleo son muy prolongados (alta duración) aunque la incidencia sea baja, o puede deberse a que hay una alta frecuencia de desempleo (alta incidencia) aunque con baja duración.

Las estimaciones anteriores nos permiten hacer inferencias sobre las tasas de incidencia del desempleo por grupos de la población y de esa forma contribuir a la caracterización del desempleo de los mismos grupos. La incidencia se

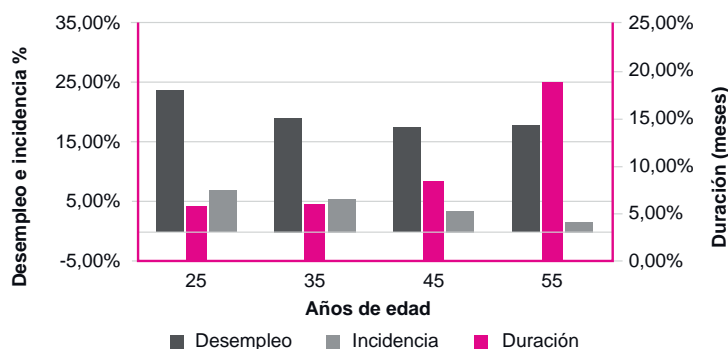
puede estimar como la razón de la tasa de desempleo dividida por la duración media de desempleo. Dichas estimaciones se presentan en la tabla 12 a continuación.

Tabla 10. Hombres.

Jefe de Hogar=0, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.228	0.249	0.266	0.276	0.191	0.202	0.215	0.223	0.173	0.184	0.193	0.198	0.172	0.183	0.191	0.200
Modelo 2	0.287	0.300	0.312	0.319	0.264	0.274	0.286	0.293	0.257	0.268	0.279	0.286	0.266	0.279	0.290	0.297
Jefe de Hogar=0, Estado Civil=1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.237	0.247	0.260	0.271	0.189	0.199	0.212	0.220	0.171	0.179	0.191	0.199	0.179	0.185	0.195	0.201
Modelo 2	0.232	0.242	0.253	0.260	0.213	0.222	0.231	0.238	0.208	0.219	0.228	0.234	0.218	0.229	0.238	0.245
Jefe de Hogar=1, Estado Civil =0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.131	0.215	0.224	0.232	0.161	0.173	0.178	0.182	0.143	0.156	0.161	0.163	0.152	0.157	0.165	0.167
Modelo 2	0.183	0.220	0.229	0.236	0.184	0.200	0.204	0.208	0.175	0.194	0.199	0.200	0.194	0.202	0.212	0.212
Jefe de Hogar=1, Estado Civil = 1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.221	0.231	0.242	0.253	0.182	0.187	0.197	0.206	0.167	0.174	0.181	0.186	0.171	0.178	0.186	0.191
Modelo 2	0.174	0.182	0.191	0.199	0.160	0.168	0.175	0.181	0.157	0.166	0.173	0.178	0.166	0.175	0.183	0.188

Fuente: elaboración propia

Gráfico 7. Incidencia, desempleo y duración de hombres casados jefes de hogar.



Fuente: elaboración propia

Las estimaciones indican que la incidencia disminuye con la edad y con los niveles de educación. Es decir, a medida que aumenta la edad de las personas, la frecuencia con la que quedan desempleados disminuye. Lo mismo ocurre a medida que aumentan los niveles de educación. A partir de la información presentada se puede caracterizar el desempleo de los diferentes grupos.

En general, la conclusión que se obtiene es que la duración del desempleo aumenta con la edad y con

Tabla 11. Mujeres.

Jefe de Hogar = 0, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.335	0.354	0.349	0.322	0.276	0.291	0.285	0.262	0.229	0.243	0.239	0.217	0.198	0.209	0.205	0.187
Modelo 2	0.350	0.368	0.359	0.328	0.289	0.303	0.297	0.272	0.242	0.256	0.252	0.227	0.211	0.222	0.218	0.197
Jefe de Hogar=0, Estado Civil = 1																
Edad	25	35	45	55												
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.338	0.348	0.344	0.321	0.272	0.285	0.280	0.261	0.225	0.237	0.234	0.216	0.195	0.204	0.200	0.185
Modelo 2	0.338	0.346	0.341	0.317	0.271	0.283	0.279	0.260	0.226	0.237	0.233	0.215	0.196	0.205	0.201	0.185
Jefe de Hogar=1, Estado Civil =0																
Edad	25	35	45	55												
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.310	0.334	0.336	0.315	0.264	0.281	0.275	0.254	0.230	0.243	0.239	0.215	0.196	0.209	0.207	0.188
Modelo 2	0.316	0.333	0.332	0.309	0.263	0.280	0.274	0.253	0.232	0.243	0.240	0.215	0.197	0.210	0.208	0.189
Jefe de Hogar=1, Estado Civil = 1																
Edad	25	35	45	55												
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.318	0.350	0.349	0.328	0.276	0.290	0.286	0.266	0.228	0.243	0.241	0.223	0.198	0.209	0.207	0.191
Modelo 2	0.301	0.336	0.333	0.310	0.263	0.277	0.273	0.254	0.217	0.232	0.229	0.213	0.188	0.200	0.198	0.182

Fuente: elaboración propia

los niveles educativos, mientras que la incidencia disminuye. En otras palabras, el desempleo de los jóvenes es un desempleo de corta duración, pero las instancias de desempleo son muy frecuentes. Esto es consistente con situaciones en las que hay poca experiencia laboral y poco compromiso con el mercado laboral.

Por el contrario, el desempleo de las personas mayores es de baja incidencia, pero de larga duración. Esto puede ser consistente con situaciones de obsolescencia de

las competencias (por ejemplo por cambios tecnológicos), con situaciones de discriminación y con otras que hacen difícil para las personas obtener ofertas de empleo razonables (altos salarios de reserva).

Los gráficos siguientes, contruidos a partir de la información presentada, ilustran estos puntos. Como se puede ver, la tasa de incidencia disminuye a medida que aumenta la edad, mientras que la duración del desempleo aumenta. En el caso de los hombres jefes de hogar, la incidencia de los de 55

años de edad es un tercio de la de los que tienen 25 años, pero la duración es tres veces mayor. Algo similar ocurre por niveles educativos. Los trabajadores con bajos niveles quedan desempleados con mucha frecuencia, pero duran desempleados poco tiempo. Los de altos niveles educativos quedan desempleados con baja frecuencia, pero sus procesos de búsqueda son mucho más largos.

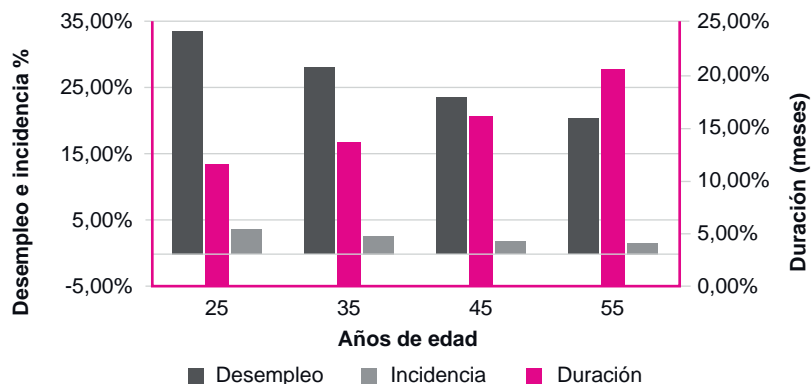
Lo interesante del resultado obtenido es que las políticas necesarias para combatir el desempleo de

Tabla 12. Tasa de incidencia del desempleo.

Tasas de incidencia del desempleo									
Hombres					Mujeres				
Años	Años	No-jefes de hogar		Jefes de hogar		No-jefes de hogar		Jefes de hogar	
Edad	Educ	Solteros	Casados	Solteros	Casados	Solteras	Casadas	Solteras	Casadas
25	0	10,36%	12,47%	31,83%	13,81%	5,00%	5,04%	5,02%	4,75%
	5	7,55%	7,97%	6,94%	8,25%	3,87%	3,57%	3,47%	3,65%
	11	3,75%	3,82%	3,80%	4,17%	2,70%	2,53%	2,53%	2,62%
	16	1,70%	1,74%	1,76%	1,96%	1,99%	1,89%	1,92%	2,01%
35	0	9,55%	10,50%	8,94%	10,71%	3,71%	3,39%	3,33%	3,37%
	5	6,73%	6,86%	6,41%	6,93%	2,71%	2,86%	2,52%	2,57%
	11	3,36%	3,37%	3,36%	3,52%	1,92%	3,49%	1,79%	1,86%
	16	1,45%	1,43%	1,53%	1,43%	1,44%	2,68%	1,38%	1,41%
45	0	6,65%	6,84%	5,72%	6,96%	2,44%	2,35%	2,42%	2,35%
	5	3,91%	4,37%	3,90%	4,58%	1,87%	1,73%	1,80%	1,79%
	11	2,22%	2,20%	2,15%	2,26%	1,34%	1,23%	1,30%	1,29%
	16	1,02%	1,03%	1,00%	0,96%	0,99%	0,93%	0,96%	0,99%
55	0	3,51%	3,58%	3,10%	3,42%	1,73%	1,62%	1,63%	1,64%
	5	2,23%	2,26%	2,28%	2,23%	1,30%	1,19%	1,23%	1,23%
	11	1,17%	1,12%	1,11%	1,17%	0,93%	0,85%	0,89%	0,89%
	16	0,56%	0,45%	0,58%	0,41%	0,68%	0,64%	0,67%	0,65%
Promedios simples por años de edad									
25		5,84%	6,50%	11,08%	7,05%	3,39%	3,26%	3,23%	3,26%
35		5,27%	5,54%	5,06%	5,65%	2,44%	3,10%	2,26%	2,30%
45		3,45%	3,61%	3,19%	3,69%	1,66%	1,56%	1,62%	1,60%
55		1,87%	1,85%	1,77%	1,81%	1,16%	1,08%	1,10%	1,10%
Promedios simples por años de educación									
	0	7,52%	8,35%	12,40%	8,72%	3,22%	3,10%	3,10%	3,02%
	5	5,11%	5,36%	4,88%	5,49%	2,44%	2,34%	2,25%	2,31%
	11	2,62%	2,63%	2,60%	2,78%	1,72%	2,02%	1,63%	1,66%
	16	1,18%	1,16%	1,22%	1,19%	1,28%	1,53%	1,23%	1,27%

Fuente: elaboración propia

Gráfico 8. Incidencia, desempleo y duración de mujeres casadas jefes de hogar.




Fuente: elaboración propia

corta duración y de larga duración son diferentes. Generalmente, el desempleo de larga duración requiere procesos de rentrenamiento y, en el caso de los trabajadores altamente calificados, de asistencia con información especializada sobre los nichos del mercado laboral relevantes. El desempleo de corta duración muchas veces puede ser disminuido con políticas activas de generación de empleo.

Bibliografía

- Allison, P. D. (2004). *Survival Analysis Using SAS, A Practical Guide* (7th Edition). SAS Publishing..
- Baltagi, B. H. (2001). *A Companion to Theoretical Econometrics*. Blackwell Publishers.
- Cameron, A. y Trivedi, P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Cleves, M. A. (2008). *An Introduction to Survival Analysis Using Stata*. A Stata Press Publication.
- Castellar, C. y Uribe, J. I. (2003). Determinantes de la duración del desempleo en el Área Metropolitana de Cali, 1988-2000. *Archivo de Economía*, núm. 218. Departamento Nacional de Planeación.
- Ehrenberg, R. y Oaxaca, R. (1976). Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain. *AER*.
- Evans, M., Hastings, N., Peacock B. y Forbes, C. (2011). *Statistical Distributions*, Fourth Edition, Wiley.
- Fitzenberger, B. y Wilke, Ralf A. (2005). Using Quantile Regression for Duration Analysis. *ZEW Discussion Paper, 05-58, Mannheim*.
- Heckman, J. y Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52.
- Heij, C., de Boer, P., Franses, P. H., Kloek, T. y van Dijk, H. (2004). *Econometric Methods with Applications in Business and Economics*. Oxford University Press.
- Hosmer, D., Lemeshow, S. y May, S. (2008). *Applied Survival Analysis* (Wiley Series in Probability and Statistics, 2th Edition).
- Klein, J. P. y Moeschberger, M. L. (2003). *Survival for Censored and Truncated Data* (Springer, 2th Edition).
- Koenker, R. y Hallock. (2001). *Quantile regression. Journal of Economic Perspectives*, 15(4), 143-156.
- Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica*, 47(4): 939-956.
- Maddala, G. S. (1997). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.
- Martínez, H. (2003). Cuánto duran los colombianos en el desempleo y en el empleo: un análisis de supervivencia. *Archivos de Economía* 236.
- Meyer, B. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, 58(4): 757-782.
- Núñez J. y Bernal, R. (1998). El desempleo en Colombia: tasa natural, desempleo cíclico y estructural y la duración del desempleo (1976-1998), DNP, Archivos de Macroeconomía, Documento núm. 97. Bogotá, D. C.
- Pesaran, M. H. y Schmidt, P. (1999). *Handbook of Applied Econometrics*, (Vol. II: Microeconometrics). Blackwell Handbooks in Economics.
- Pollmann-Schult, M. y Büchel, F. (2005). Unemployment Benefits, Unemployment Duration and Subsequent Job Quality: Evidence from West Germany. *Acta Sociologica*, 48(1): 21-39.
- Tenjo, J. y Ribero, R. (1998). Participación, desempleo y mercados laborales en Colombia, DNP. *Archivos de Macroeconomía*, documento 81. Bogotá, D. C.
- van den Berg, G., Lindeboom, M. y Dolton, P. (2006). *Survey Non-Response and the Duration of Unemployment*. *Royal Statistical Society*.
- Viáfara C. y Uribe, I. (2008). Duración del desempleo y canales de búsqueda de empleo en Colombia 2006. *Revista de Economía Institucional*, 11(21), 2009. Universidad Externado de Colombia.
- Wichert, L. y Wilke, R. (2008). Simple non-parametric estimators for unemployment duration analysis. *Royal Statistical Society*, 57, Part 1: 117-126.
- Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.



Una revisión a los procesos de aseguramiento de la calidad en operaciones estadísticas por muestreo

Leonardo Trujillo Oyola
Luz Mery González García

Una revisión a los procesos de aseguramiento de la calidad en operaciones estadísticas por muestreo

Leonardo Trujillo Oyola: Estadístico de la Universidad Nacional de Colombia, magíster en Estadística y doctor en Estadística Social de la Universidad de Southhampton, Inglaterra. Profesor del Departamento de Estadística de la Universidad Nacional de Colombia.
Correo electrónico: ltrujillo@unal.edu.co

Luz Mery González García: Estadística de la Universidad Nacional de Colombia, magíster en estadística y doctora en Estadística de la Universidad de San Pablo, Brasil. Profesora del Departamento de Estadística de la Universidad Nacional de Colombia. Correo electrónico: lgonzalezg@unal.edu.co

Fecha de recepción: 21 de junio de 2013
Fecha de aceptación: 21 de octubre de 2013

Resumen: este artículo presenta una introducción a la historia de las operaciones estadísticas por muestreo estadístico así como una revisión de los procesos de aseguramiento de la calidad de las mismas a nivel internacional. El enfoque está restringido al aseguramiento de la calidad de las encuestas por muestreo. Después de hacer una descripción histórica de las encuestas por muestreo se describen los desarrollos históricos en diversos países para garantizar el aseguramiento de la calidad en operaciones estadísticas por muestreo. Finalmente, se hacen algunas recomendaciones a los lineamientos del proceso actual de aseguramiento de la calidad en operaciones estadísticas por muestreo en Colombia, actividad liderada en el momento por el Departamento Administrativo Nacional de Estadística (DANE).

Palabras clave: calidad en encuestas, historia del muestreo, inferencia basada en el diseño, inferencia basada en el modelo, muestreo no probabilístico, muestreo probabilístico.

Abstract: this article presents an introduction to the history of statistics operations by statistical sampling and a review of the processes of quality assurance of them in an international level. The approach is restricted to the quality assurance of the sample surveys. After making a historical description of sample surveys, we described historical developments in different countries to ensure the quality assurance of the statistical operations by sampling. Finally, some recommendations are made to the guidelines of the current process of quality assurance in statistical operations by sampling in Colombia, activity led at the time by DANE.

Key words: surveys quality, sampling history, design-based inference, model-based inference, non-probability sampling, probability sampling.

1. Introducción

Groves et al. (2009, p. 2, sección 1.1) definen la metodología de encuestas como un campo del conocimiento, una profesión y una ciencia. Esta área del conocimiento considera la producción de operaciones estadísticas tan diferentes como los censos, las encuestas (tanto por muestreo probabilístico como no probabilístico) y los registros administrativos. Aunque esta ciencia comprende la producción de todas estas operaciones estadísticas, en este artículo nos limitaremos únicamente al aseguramiento de la calidad de encuestas por muestreo probabilístico.

La metodología de encuestas como campo del conocimiento corresponde al estudio de las fuentes de error en cualquier operación estadística y cómo hacer que los resultados producidos sean lo más precisos posible mediante la minimización de estos errores. Al final del proceso de producción, las estadísticas publicadas son usualmente estimaciones puntuales de diferentes parámetros poblacionales.

El término «operación estadística» recoge todos los procesos de conceptualización, desarrollo y aplicación de métodos y técnicas de diseño, procedimiento, recolección, procesamiento, análisis, almacenamiento, conservación y divulgación de datos estadísticos, destinados a la obtención de información estadística sobre un determinado tema de la realidad nacional. El término representa una sistematización simplificada de actividades estadísticas.

Las estimaciones son el resultado final de una serie de operaciones o procesos que comienzan con la definición de los objetivos, el uso de conceptos o definiciones estandarizadas, la definición de una población objetivo, la escogencia de un marco muestral, la escogencia

de un diseño de muestreo, el entrenamiento y supervisión de entrevistadores y codificadores, la edición, imputación y estimación con los ajustes apropiados por ausencia de respuesta. La calidad de cada uno de estos procesos influencia la calidad de la estimación finalmente producida y publicada. La estimación por sí misma –el número publicado– no dice nada acerca de esta calidad o las diferentes condiciones bajo las cuales fue obtenida. Por lo tanto, los efectos de los diferentes procesos dentro de una operación estadística se encuentran desapercibidos tras este número o estimación.

Tanto buenas estadísticas como malas estadísticas lucen esencialmente iguales en su número. Por esta razón, es justo que el usuario esté informado de las características esenciales de la operación estadística, incluyendo sus fortalezas y debilidades. Esto implica que el productor de la operación estadística tiene la obligación de describir las características esenciales de la misma. De esta manera, un usuario puede evaluar e interpretar los resultados de una encuesta desde sus propios objetivos particulares. Los usuarios podrían estar interesados en conocer los detalles de la metodología estadística, tales como la escogencia de estimadores particulares y de su varianza. Es por esto que la información al usuario debería incluir: a) una descripción precisa de la población y los dominios de la población acerca de los cuales se pueden obtener inferencias estadísticas válidas, b) definiciones de las variables de estudio e información de en cuáles de estas definiciones fueron incluidas como parte de los ítems del cuestionario, c) por lo menos estimaciones insesgadas del error cuadrático medio y sus componentes: la varianza y el sesgo en el caso de operaciones por muestreo probabilístico.

En una encuesta por muestreo, por ejemplo, la varianza tiene di-

versas fuentes de origen tales como el error muestral, el error de respuesta, el error de ausencia de respuesta, datos faltantes y errores de procesamiento debidos a edición e imputación. Es difícil identificar y medir apropiadamente todas estas fuentes de error. Estimaciones insesgadas de la varianza total pueden ser obtenidas a un gran costo. La cantidad de recursos requeridos para medir la calidad de los datos, en una operación estadística, normalmente riñe con restricciones de presupuestos, recursos humanos y tecnológicos.

Las entidades responsables de diversas operaciones estadísticas en un país pueden no siempre vivir bajo el ideal representado por los puntos a) a c) y en particular el punto c) puede causar dificultades. Sin embargo, todas estas instituciones están seriamente comprometidas a entregar los mejores indicadores posibles de su operación estadística. En la actualidad, esto se rige bajo algunos estándares internacionales plasmados en el Código de Buenas Prácticas (véase, por ejemplo, DANE 2011a).

El objetivo de este artículo es presentar los lineamientos que se deberían tener en cuenta para el aseguramiento de la calidad de operaciones estadísticas por muestreo probabilístico y hacer una comparación con los estándares que son tenidos en cuenta por diversas oficinas de estadística alrededor del mundo. Esto puede ser un insumo para la consolidación de los criterios que actualmente son tenidos en cuenta para el aseguramiento de la calidad de operaciones estadísticas por muestreo probabilístico en el DANE así como para otras instituciones productoras de estadística a nivel nacional.

El artículo se encuentra organizado como sigue: la sección 1 corresponde a la presente introducción; la sección 2 hace una definición

de los tipos de operaciones estadísticas entre censos, encuestas y registros administrativos; la sección 3 presenta una breve historia e introducción a la filosofía de las encuestas por muestreo; la sección 4 presenta los lineamientos o estándares de calidad internacionales para operaciones por muestreo y su variación histórica; la sección 5 se enfoca a los estándares actualmente considerados por el DANE en Colombia, y la sección 6 presenta algunas conclusiones de esta revisión así como algunas recomendaciones a futuro.

2. Tipos de operaciones estadísticas

Esencialmente hay tres tipos de operaciones estadísticas: censos, encuestas y registros administrativos. Con el fin de hacer una distinción entre estos tipos de operaciones y enfocarnos específicamente en las operaciones estadísticas correspondientes a encuestas por muestreo probabilístico, haremos una definición de cada una.

Los *censos* son generalmente desarrollados e implementados por el gobierno de un país. Constituyen esfuerzos sistemáticos para contar una población completa, en algunos países con periodicidades establecidas de uno cada diez años, como en los Estados Unidos. Los censos son útiles puesto que constituyen un núcleo de datos estadísticos oficiales mediante el cual otros análisis demográficos, estimaciones por encuestas y registros administrativos pueden ser calibrados (Nirel y Glickman, 2009, Pfeffermann y Rao, 2009). Un censo poblacional ha sido definido recientemente como «una operación que produce a intervalos regulares los conteos oficiales (o *benchmarks*) de la población en el territorio de un

país y en sus más pequeños subteritorios geográficos, junto con la información de un número selecto de características sociales y demográficas de la población entera» (United Nations Economic Commission for Europe, UNECE, 2006, p. 6).

Una *encuesta por muestreo* es un método sistemático para recoger información de ciertos individuos (en una muestra) con los propósitos de construir descriptores cuantitativos de los atributos de una población más grande de la cual los individuos son miembros. La frase «en una muestra» aparece en paréntesis, pues algunas veces las encuestas intentan medir todos los individuos en una población y algunas veces solo en una muestra (un subconjunto de la población). Es decir, la recolección de la información en toda la población o censo se puede considerar como un tipo de encuesta. Aunque la teoría de la probabilidad fue establecida en el siglo XVIII (Jeffrey, 1992), su aplicación en el campo del muestreo de encuestas solo comenzó hasta el siglo XX (Smith, 1978). Las primeras aplicaciones fueron tomar «1 de N» selecciones sistemáticas de los resultados de un censo. Estas fueron *encuestas por muestreo probabilístico*; es decir, aquellas donde cada registro tenía una probabilidad conocida diferente de cero de ser seleccionado en la muestra. La aplicación en muestreos agrícolas conllevó a la aplicación de encuestas por muestreo probabilístico de áreas, y luego la misma técnica fue extendida a muestras de hogares en ciudades (Hansen y Hauser, 1945). Las encuestas para las cuales algunos de los individuos de la población tienen probabilidad cero de ser seleccionados se consideran como encuestas por *muestreo no probabilístico*.

Los *registros administrativos* (Sanz, 2006) consideran la recolección de información referida a una unidad individual que ha sido diseñada, recogida y mantenida por una insti-

tución con el propósito de controlar y tomar decisiones relativas a esa unidad individual. Las acciones a tomar incluyen funciones como permitir, registrar, inspeccionar, asegurar, regular, revisar, diagnosticar, tratar, cobrar, pagar, o proporcionar otros beneficios o penalizaciones. La diferencia de un registro administrativo con los censos y las encuestas es que los dos últimos se diseñan con el fin de controlar y tomar decisiones relativas a la población objetivo, no al individuo en sí. Adicionalmente, en los censos y en las encuestas, el tiempo de recolección es un proceso finito mientras que en los registros administrativos el tiempo de recolección es permanente.

3. Historia y filosofía del muestreo de encuestas

El uso de muestras para estimar la estructura de una población objetivo ha sido común desde el comienzo de la humanidad. Sin embargo, el muestreo como un método científico es un fenómeno mucho más reciente. El primer uso del muestreo estadístico del que se tenga conocimiento data del trabajo de A.N. Kiaer en la Oficina de Estadísticas de Noruega hacia finales de la década 1890. En 1895, en la reunión del International Statistical Institute (ISI) en Berna (Suiza), Kiaer presentó una serie de artículos que demostraban empíricamente la precisión de estimadores de totales y promedios de población finita, los cuales podían ser obtenidos a través de muestras relativamente pequeñas (Kiaer 1895, 1897, 1901, 1997).

El legado de Kiaer en lo que él llamó «el método representativo de encuestas estadísticas» encontró oposición en algunos de los miembros más prominentes del ISI en aquella época. La mayoría de ellos eran estadísticos del gobier-

no, quienes no tenían nada que ver con menos que un cubrimiento total de cualquier proceso de recolección oficial.

Kiaer nunca definió precisamente «el método representativo» en sus escritos. Sin embargo, él creía en un proceso de seleccionar cuidadosamente unidades representativas en diferentes niveles de desagregación de la población de estudio. También insistió en tamaños de muestra sustanciales en todos los niveles del proceso de selección y se aseguraba de la representación de cada uno de los subgrupos (Seng, 1951; Smith, 1978).

A pesar de ser contrariado, las ideas de Kiaer finalmente ganaron reconocimiento y en 1903 el ISI dio validez al «método representativo» como un método científico de recolección de datos. El siguiente paso significativo en el desarrollo de la teoría de muestreo vino con el trabajo de Bowley (1906). En su charla ante la sección de estadística y ciencia económica de la Asociación Británica para el Avance de la Ciencia, Bowley puso en consideración la combinación del uso de muestras con la inferencia estadística, presentando por primera vez un método que permitía asegurar la precisión de estimadores obtenidos por muestras aleatorias. Incluso desarrolló fórmulas para la varianza de los estimadores derivados desde una muestra aleatoria estratificada.

Bowley (1913) realizó un estudio de pobreza que tuvo en cuenta uno de veinte hogares en una muestra sistemática de cuatro ciudades inglesas escogidas a propósito. Allí se discuten por primera vez los conceptos de medición y de errores de muestreo en el contexto de un muestreo probabilístico.

Hacia 1920, la difusión del muestreo era amplia y la gran pregunta era cómo aplicar el «método

representativo». El argumento se encontraba entre dos técnicas básicas de selección:

- *Selección aleatoria*: donde cada unidad tenía la misma probabilidad de inclusión diferente de cero en la muestra.
- *Selección por conveniencia*: donde la muestra se escogía de tal manera que fuese un espejo de las características de la población (por ejemplo, un muestreo de cuotas).

El muestreo estratificado fue considerado como una extensión del método de selección aleatoria dado que los estratos eran muestreados en proporción a su representación en la población. En la práctica las muestras aleatorias eran extraídas usando un muestreo sistemático de una lista ordenada con un arranque aleatorio. Muestras que fueran verdaderamente aleatorias simples solo se hicieron realidad hacia finales de la década de 1920 con la primera publicación de tablas de números aleatorios. En 1924, el ISI estableció un comité para estudiar «La aplicación del método representativo en estadística», y en 1925 este comité reportó que las dos técnicas de selección eran válidas para obtener «muestras representativas». También recomendó que cualquier uso del método debería permitir la estimación de «en qué magnitud el error de las estimaciones de la encuesta eran aceptables» y que el reporte final debería presentar no solo los resultados de la encuesta sino también «una reseña explícita en detalle del método para seleccionar la muestra».

El próximo y tal vez más significativo paso en el desarrollo de la teoría moderna del muestreo fue el artículo de Jerzy Neyman (1934). Este artículo contiene problemáticamente la primera derivación rigurosa del concepto de un intervalo de confianza. Sin embargo, el artículo

lo presenta una detallada crítica del muestreo por conveniencia y la introducción y refinamiento de lo que se conoce como la aleatorización o la aproximación «basada en el diseño» en el muestreo de encuestas. No fue más necesario ni relevante hablar acerca de «muestras representativas» puesto que la representatividad era ahora una propiedad del método de muestreo y solo aquellos métodos de muestreo que podían generar probabilidades de inclusión (es decir, muestras probabilísticas con selección aleatoria) eran representativos puesto que solo ellas podían generar intervalos con las propiedades de confianza requeridas por Neyman (Kruskal and Mosteller, 1979a, 1979b, 1979c y 1980).

El artículo de Neyman fue tan influyente que aproximadamente durante los quince años siguientes (1934-1949) la aleatorización se convirtió en el paradigma dominante de la inferencia por muestreo. El muestreo por conveniencia no fue más considerado como un método científico de recolección de datos aunque siguió siendo utilizado ampliamente en investigación de mercados y encuestas de opinión.

Explícitamente en el artículo de Neyman se encontraba la idea de que las muestras aleatorias con probabilidades de inclusión desiguales no eran posibles pero deseables desde un punto de vista de eficiencia. Hansen y Hurwitz (1943) extendieron esta idea cuando introdujeron el muestreo con probabilidades proporcionales al tamaño como una manera efectiva de seleccionar muestras multietápicas. Cochran (1942, 1953) desarrolló la teoría del estimador de razón y el estimador de regresión usando por primera vez argumentos basados en un modelo probabilístico para la población objetivo. Mahalanobis (1946) desarrolló procedimientos para mezclar muestras que llevaron a las ideas del muestreo repetido y técnicas

como *bootstrap* y *jackknife*. Paterson (1950) desarrolló la teoría para el diseño y análisis de encuestas repetidas. Horvitz y Thompson (1952) introdujeron la teoría general del muestreo con probabilidades desiguales sin reemplazo.

Brewer (1963) sugirió por primera vez una aproximación basada en el modelo para la estimación de encuestas por muestreo. En particular, Brewer notó que si se está preparado para asumir que los valores de la población objetivo siguen un modelo de regresión lineal simple con valores de la variable independiente en el modelo conocidos para todas las unidades en la población se pueden encontrar estimadores que corresponden a un predictor lineal insesgado del total poblacional con mínima varianza. La idea de Brewer no fue aceptada inmediatamente por la comunidad estadística (Brewer, 1995). Sin embargo, Royall (1970) redescubrió los resultados de Brewer, hizo una extensión de ellos y demostró que una estrategia de muestreo basada en un muestreo por conveniencia de las unidades más grandes en una población junto con un predictor de regresión lineal del total poblacional generaba un error muestral con varianza más pequeña que cualquier otra estrategia basada en un estimador lineal insesgado y una muestra del mismo tamaño.

El estado actual de la teoría de muestreo de encuestas fluye entre estas dos teorías: la inferencia basada en el diseño y la inferencia basada en el modelo. La práctica de inferencia en encuestas por muestreo está todavía atada en su mayoría a la teoría de aleatorización (inferencia basada en el diseño), principalmente por el amplio número de técnicas para la práctica del muestreo bajo esta filosofía desde los años cuarenta. Sin embargo, la teoría moderna de la inferencia en muestreo parece

moverse hacia la teoría basada en el modelo, aunque aún de manera lenta incluso en institutos nacionales de estadísticas oficiales.

4. Estándares de calidad internacionales para encuestas por muestreo probabilístico

En la presente sección discutiremos las condiciones o características que, de acuerdo con diferentes oficinas de estadísticas nacionales, se deben tener en cuenta en encuestas por muestreo. Estas condiciones o características nos permitirán definir los criterios y lineamientos a ser evaluados en los procesos de aseguramiento de la calidad.

Las principales oficinas de estadísticas nacionales (NSO, National Statistical Offices) por largo tiempo han reconocido la importancia de adherirse a algunas políticas de información que redunden en la calidad de las estadísticas producidas (Särndal, Swensson y Wretman, 1992). Las Naciones Unidas han estado al frente de la promoción de buenas prácticas con el fin de garantizar la calidad de los datos en diferentes operaciones estadísticas. En 1948, la Subcomisión de las Naciones Unidas en Muestreo Estadístico de la Comisión Estadística produjo un documento titulado «The Preparation of Sampling Survey Reports», el cual fue revisado por un grupo de estadísticos reconocidos internacionalmente (Darmois, Deming, Fisher, Mahalanobis y Yates como consultores).

En 1974, la Oficina del Censo de los Estados Unidos (United States Bureau of the Census) produjo un reporte técnico conocido como «Standards for Discussion and Presentation of Errors in Data». Una

versión revisada (González, Ogus, Shapiro y Tepping, 1975) sugiere el tipo de información sobre medición de errores que debería ser incluido en los reportes de encuestas y sugiere el tipo de errores que los usuarios deberían tener en cuenta en el momento de evaluar los resultados de las mismas.

Estos desarrollos llevaron a las NSO a preparar sus propios lineamientos para sus declaraciones de calidad. En 1982, se organizó una reunión durante la Conferencia de Estadísticos Europeos (Conference of European Statisticians, Statistical Commission and Economic Commission for Europe, United Nations Economic and Social Council) llamada «Timeliness, Cost and Quality Attributes of Statistics», con participación de las oficinas de estadística de Canadá, España, Francia, Holanda, Reino Unido, Suecia y Suiza, así como de la International Labor Organization y la Comunidad Económica Europea. La reunión basó su discusión en un reporte preparado por Statistics Sweden llamado «Some Reflections on Different Approaches to the Problems of Presenting the Quality of Statistics», el cual presentaba una comparación de los lineamientos de calidad entre Canadá, Estados Unidos y Suecia. Este reporte se comparó con el artículo de González et al. (1975) y con otro del Departamento de Comercio de los Estados Unidos (U.S. Department of Commerce, 1978a).

En el U.S. Department of Commerce (1978b), el error de perfil se ilustra para el caso particular de la *Current Population Survey*, una encuesta nacional que entrega las estadísticas laborales (empleo y desempleo) de los Estados Unidos. El término «error de perfil» se describe como la forma comprensiva y sistemática de estudiar los diferentes procesos de una operación estadística desde su diseño hasta la difusión de sus

resultados. Este permite al usuario conocer las limitaciones de las estadísticas y guiar a los productores a identificar los procesos propios que requieren de una mayor atención en cuanto a su rediseño o mejor control con el fin de mejorar la calidad de sus estimaciones.

El error de perfil para esta encuesta tuvo en cuenta los siguientes criterios:

1) *Proceso de muestreo*: marco muestral, selección de la muestra, controles de calidad del proceso de muestreo.

2) *Proceso de trabajo de campo*: procedimiento de recolección de los datos, diseño del cuestionario, personal para la recolección de los datos, capacitación de los encuestadores, control de calidad del trabajo de campo.

3) *Procesamiento de los datos*: operaciones de ingreso de datos (digitación), edición, crítica e imputación, controles de calidad al procesamiento de los datos.

4) *Proceso de estimación*: procedimientos de ponderación, procedimientos de estimación, controles de calidad al procedimiento de estimación.

5) *Análisis y difusión*.

Adicional a los criterios anteriores, la Conferencia de Estadísticos Europeos en 1982 acordó que las presentaciones de calidad deberían incluir también (Sarndal et al., 1992, p. 640):

a. Información básica de la fuente de datos. Incluyendo definiciones y clasificaciones.

b. Cobertura de los datos. La calidad del marco muestral, como un *proxy* a la población objetivo, debería incluir la revisión de duplicaciones, omisiones y problemas de definición.

c. Descripción de los métodos de selección y estimación. La variabilidad total de las estimaciones, así como los efectos del error muestral, error de ausencia de respuesta y errores de procesamiento.

d. Tasas de respuesta. Porcentaje de la muestra de la cual se obtuvo respuestas (pregunta por pregunta si es apropiado). Diferencias conocidas en las características de los respondientes y los no respondientes. Una breve indicación de los métodos de imputación o estimación usados para compensar la ausencia de respuesta.

e. Errores de muestreo. Descripción de cómo fueron calculados los errores estándar y las medidas relacionadas así como su interpretación. Si el muestreo es probabilístico, estimaciones del error estándar junto con una explicación de cómo deben ser interpretados. Si el muestreo es no probabilístico, las implicaciones de hacer inferencias con los datos de la encuesta.

f. Cambios metodológicos. Cambios significativos en los procedimientos de un periodo de la operación estadística a otro que pudieran afectar su comparabilidad a través del tiempo.

g. Comparabilidad con otras fuentes. Información de la comparabilidad con estadísticas en el mismo tema recogidas por otras fuentes a nivel local o internacional. Si existen datos similares de otras fuentes, estas se deben identificar. Cuando sea apropiado, se debe describir los conjuntos de datos que difieren y las razones de estas diferencias.

En 1987, Statistics Canada publicó sus políticas acerca del proceso de calidad de los datos de sus investigaciones en cuanto a su metodología

estadística (Statistics Canada, 1987, parágrafo 14). Todos los criterios del U.S. Department of Commerce (1978a) y de la Conferencia de Estadísticos Europeos en 1982 fueron también tenidos en cuenta por Statistics Canada, y además se consideró:

a. Comparabilidad en el tiempo: La comparabilidad de los resultados de la misma actividad a un periodo anterior de referencia determinado, especialmente si ha habido un cambio en la metodología, conceptos o definiciones. Si tal cambio afecta la comparabilidad de un periodo a otro, una estimación cuantitativa de este efecto se debe medir. Los efectos de calibraciones, *benchmarking* y revisiones en la comparabilidad sobre el tiempo, también se deben describir.

b. Otros aspectos importantes: Problemas de recolección atípicos, malas interpretaciones de los encuestados acerca de algunas definiciones, cambios en las clasificaciones usadas, entre otros.

c. Sesgo de respuesta: Mala comprensión de algunas definiciones por parte de los encuestados, problemas del cuestionario, entre otros.

d. Efecto de la edición e imputación.

e. Ajustes estacionales: Mediciones del impacto, significancia de los ajustes y una explicación de cómo se deben interpretar. Ejemplos de tales medidas son el cambio porcentual absoluto promedio de las revisiones del último año del factor estacional y la estadística MCD (*months for cyclical dominance*; es decir, meses de dominancia cíclica).

Los lineamientos de Statistics Canada consideran que se deben entregar dos niveles de documentación:

a. Reportes generales a los usuarios que estén diseñados para una audiencia amplia, con el fin de ayudarlos a la interpretación de los datos y en decidir si los datos son apropiados para sus fines particulares.

b. Reportes técnicos, a público más especializado, que son definitivos y exhaustivos y den completa y detallada información en los métodos particulares considerados en cada uno de los procesos de la operación estadística.

En 2003, el Fondo Monetario Internacional publicó lineamientos para evaluar la calidad de las estadísticas de finanzas públicas en los que incluyó condiciones previas de la calidad como: entorno jurídico e institucional; recursos humanos, informáticos y financieros; pertinencia de las estadísticas en cuanto a necesidades de los usuarios, y gestión de la calidad. Se incluyó, adicionalmente:

a. Garantías de integridad: La imparcialidad en los análisis estadísticos, la selección de las fuentes de información y la información divulgada son solo debidas a consideraciones estadísticas, precaución en la interpretación de los datos por externos, divulgación sobre las condiciones de recopilación de los datos, información al público en general sobre el acceso a la información por parte de autoridades del gobierno antes de ser divulgada y normas éticas, entre otros.

b. Rigor metodológico: Conceptos, definiciones, alcance, clasificaciones, nomenclatura, y registro y valoración de flujos y saldos están ajustados a directrices internacionales.

c. Exactitud y fiabilidad: Programas de recopilación son adecuados, flexibles a la hora de in-

corporar cambios, se ajustan a las definiciones, al alcance, a la clasificación y a la valoración. El sistema de recopilación permite proveer datos oportunos, y hay seguimiento de la exactitud de los datos declarados.

d. Oportunidad del dato: La periodicidad, la puntualidad, la coherencia y los procesos de revisión del dato son acordes con las normas de divulgación.

e. Acceso: El usuario tiene acceso al microdato y al metadato pertinente y actualizado.

En el caso particular de los índices que son basados en datos recogidos a través de encuestas específicas o registros administrativos o una combinación de ellos, los mismos lineamientos aplican. Esto aplica a los datos sobre la variable principal (precios en el caso de índices de precios y cantidades en el caso de índices de volumen) y sus ponderaciones. Una atención especial debe darse en este caso a aspectos conceptuales y metodológicos en la fórmula de cálculo. Su apropiada descripción con una definición precisa de los conceptos económicos que se intentan medir así como de la metodología empleada (fórmula del índice, sistema de ponderación, cálculo del índice a varios niveles de desagregación, entre otros).

Otras experiencias recientes en otros países son las realizadas por Eurostat (2007), UK Statistics Authority (2009, Reino Unido), INEGI (2010, México), Statistics South Africa (2010), OECD (2011), entre muchos otros. Para más referencias internacionales, Araque (2013) hace una revisión de la temática en torno a las buenas prácticas del Sistema Estadístico Europeo así como de los sistemas estadísticos nacionales de países como Alemania, Canadá, España y Francia.

Groves et al. (2009, p. 48) desgrega el error de perfil considerando los siguientes dos aspectos de cualquier encuesta: la medición de constructos y la descripción de atributos poblacionales. Estos dos aspectos cubren los lineamientos que hasta ahora han sido considerados por diferentes oficinas de estadística en esta sección.

El diagrama 1 muestra estas dos dimensiones. Las inferencias tanto en medición como en representación van de lo abstracto a lo concreto. La dimensión de medición a la izquierda describe cuáles datos serán recogidos acerca de las unidades observacionales en la encuesta mientras la dimensión de representación a la derecha se refiere a qué poblaciones están siendo descritas por la misma.

En cuanto a la dimensión de medición, los constructos son los elementos de información que desea obtener el productor de la operación. Su naturaleza es abstracta. Las mediciones son formas de obtener la información acerca de un constructo. La diferencia entre mediciones y constructos se conocen en la teoría de la medición en psicometría como *validez*.

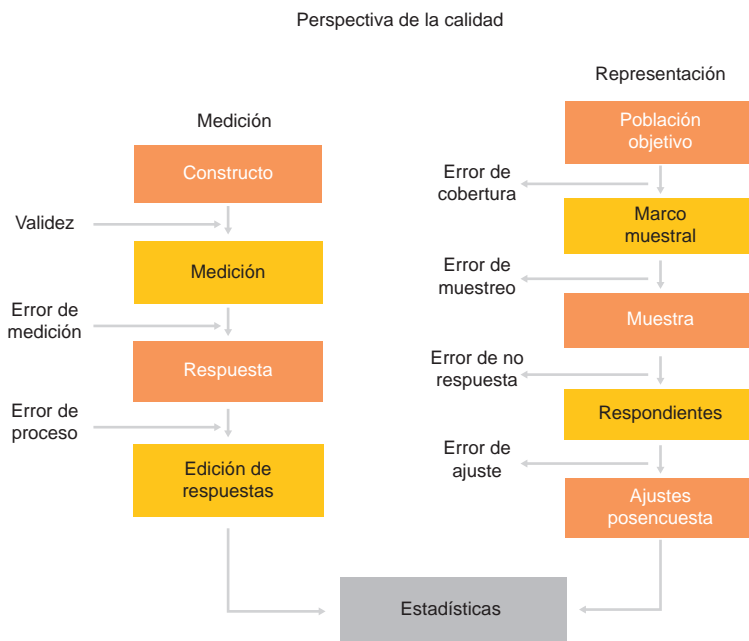
Cuando se utilizan preguntas como el mecanismo de medición, los encuestados pueden usar una variedad de medios para producir una respuesta (pueden utilizar su memoria, pueden recurrir a archivos, pueden pedir ayuda a otra persona, entre otros). La diferencia entre la respuesta y la medición se conoce como *error de medición o sesgo*.

En algunos métodos de recolección, la medición inicial pasa por una serie de revisiones como chequeos de rangos de respuestas, valores que estén por fuera de límites predeterminados o simplemente la búsqueda de patrones atípicos de respuesta u *outliers*. Las respuestas corregidas o editadas según

estas reglas de validación son las consideradas finalmente para cada individuo. Los posibles errores en

este paso de revisión se conocen como *errores de procesamiento o de proceso*.

Diagrama 1. Desagregación del error de perfil en una encuesta por muestreo probabilístico.



Fuente: Groves et al. (2009).

La siguiente dimensión en el diagrama 1 corresponde a la dimensión de representación. El gran cambio entre ambas dimensiones es que la calidad será medida sobre las estadísticas en general, no sobre los valores o respuestas individuales. En esta segunda dimensión, la primera caja describe la población objetivo o conjunto de unidades a ser estudiado. La población muestreada o en el marco muestral corresponde al conjunto de miembros de la población objetivo que tienen alguna probabilidad mayor de cero de ser seleccionados en la muestra o simplemente de ser encuestados. El marco muestral corresponde a una lista de todas las unidades en la población objetivo. Sin embargo, algunas veces es un conjunto de unidades imperfectamente relacionado con los miembros de la población y entonces puede tener problemas de subcobertura debido a elementos faltantes en el marco; o a la apa-

rición de unidades no elegibles en el marco por no ser miembros de la población objetivo (sobrecobertura). Estos dos inconvenientes definen el *sesgo o error de cobertura*.

Luego, una muestra es seleccionada del marco muestral. Esta muestra es el conjunto del cual se tomará la información. Por razones de costo o logística, no todas las personas en el marco muestral son medidas y algunas son ignoradas. Esta no observación de algunas unidades introduce desviación de las estadísticas producidas en la muestra y las estadísticas que hubieran sido obtenidas con la información del marco muestral completo. El *error de muestreo* está representado entonces por el sesgo muestral y la varianza muestral.

En casi todas las operaciones estadísticas, el intento de obtener la medición de los individuos en la

muestra no siempre se consigue. Los que son medidos se conocen como respondientes y su complemento como no respondientes. En algunos casos, definir si un individuo es respondiente o no puede ser difícil puesto que algunos solo brindan información parcial de la totalidad de mediciones. El *error de ausencia de respuesta* aparece cuando los valores obtenidos en los respondientes difieren de los que se hubieran obtenido con los datos de la muestra completa. Estos errores se conocen como sesgos de ausencia de respuesta.

Por último, cuando se tiene un conjunto de datos para los respondientes algunos valores faltantes son reemplazados mediante procedimientos de imputación o los estimadores finales son ponderados para tener en cuenta la ausencia de respuesta. Estos ajustes se conocen como ajustes posencuesta y en caso de no ser correctamente aplicados producirían errores de ajuste.

Una buena política de aseguramiento de la calidad debe buscar maneras de minimizar tanto el tamaño como el impacto de todos los errores presentados en el diagrama 1.

5. Estándares de calidad actuales en Colombia

La medición del error en encuestas por muestreo probabilístico ha sido abordada por oficinas de estadística tanto a nivel nacional como a nivel mundial. En particular, en Colombia, el DANE se ha encargado de direccionar y promover el aseguramiento de la calidad en los procesos estadísticos, trabajo que se ha evidenciado en especial desde el año 2006, cuando se inició con los procesos de certificación de sus operaciones estadísticas como de otras entidades.

El Departamento Administrativo Nacional de Estadística (DANE), en su función como coordinador del Sistema Estadístico Nacional (SEN) y en el marco del proyecto de Planificación y Armonización Estadística, viene desarrollando un proceso de certificación de calidad de operaciones estadísticas desde el año 2006 en el cual se han emitido más de 100 conceptos de evaluación (2006-2008: 42 evaluaciones, 2010: 10 evaluaciones, 2011: 20 evaluaciones, 2012: 40 evaluaciones, 2013: 60 evaluaciones a ser desarrolladas). En este sistema se evalúan

tanto censos, índices, muestras probabilísticas y no probabilísticas como registros administrativos.

El proceso inicia a través de la recolección de la información sobre las principales características de una operación estadística en un formulario de características técnicas recogido directamente en la entidad a ser evaluada. Usando esta información recolectada en el formulario, la evaluación se hace a través de una Comisión de Expertos Independientes externa al DANE, a través de instrumentos predefinidos de calificación

como matrices de calificación, reuniones con responsables y reuniones con usuarios. La calificación final concluye con una certificación de la operación en tres posibles categorías: A, B o C o la negación a ser certificada. Finalmente, se define con los responsables acerca de un plan de mejora con actividades y fechas límites que permitan considerar los cambios sugeridos por la Comisión de Expertos. Toda la información detallada sobre el proceso de evaluación de la calidad en el DANE puede obtenerse en DANE, 2011b.

Conclusiones

La medición del error en encuestas por muestreo probabilístico ha sido abordada por oficinas de estadística tanto a nivel nacional como a nivel mundial. En particular, en Colombia, el DANE se ha encargado de direccionar y promover el aseguramiento de la calidad en los procesos estadísticos, trabajo que se ha evidenciado en especial desde el año 2006, cuando se inició con los procesos de certificación de sus operaciones estadísticas como de otras entidades.

Los estándares de calidad en encuestas por muestreo probabilístico van desde la concepción de la idea, el constructo, hasta la divulgación del dato, cuidando en cada paso no solo el control de las fuentes de error (validez, error de medición, error de proceso, error de cobertura, error de muestreo, error de ausencia de respuesta, error de ajuste, etc.), sino también su debida documentación y divulgación, siempre teniendo en cuenta el por qué, el para qué, el dónde, el cuándo, la fuente y el costo.

En el caso particular de las encuestas por muestreo probabilístico se

hace importante evaluar, adicional a los demás criterios comunes con los censos y los registros administrativos (marco normativo, recursos, objetivo de la investigación, pertinencia de la fuente, diseño de cuestionarios, capacitación de los encuestadores, bases de datos, cobertura de los datos, recolección, crítica, codificación), otros criterios particulares (entre estos, actualizaciones del marco muestral, selección de la muestra, procedimientos de estimación, ajustes por ausencia de respuesta, imputación, análisis y publicación de medidas de precisión). Cada uno de estos criterios y su pertinencia han sido estudiados en el desarrollo de este artículo.

Es importante resaltar el esfuerzo que hace el DANE no solo por evaluar los procesos estadísticos de otras entidades sino también de autoevaluarse a sí mismo a través de un Comité de Expertos Independientes.

Un aspecto a considerar en discusiones futuras es la posible subjetividad que pueden tener los evaluadores que hacen que diferentes evaluaciones a cargo de diferentes comités podrían hacer sobre una

misma operación. Este fue un tipo de error no tenido en cuenta en este artículo, que se podría denominar como *sesgo del evaluador*. Sería interesante, en trabajos futuros, ahondar en cómo estimar este posible sesgo y las maneras de prevenirlo. Para esto se requeriría de una realización no interrumpida de evaluaciones con el fin de tener los históricos de evaluación de una misma operación estadística.

Todos los esfuerzos de evaluación presentados en este artículo conllevan a una mejora de la información, a una mayor satisfacción de los usuarios y a una mayor relevancia de la información producida por las entidades, así como de confianza y transparencia en las instituciones. Finalmente, la estimación por sí sola —el número publicado— no dice nada acerca de su calidad o las diferentes condiciones bajo las cuales fue obtenida. Los procesos de aseguramiento de la calidad aquí presentados por lo menos permiten una aproximación al verdadero valor de calidad de una operación estadística que está siendo evaluada.

Bibliografía

- Araque N. (2013). Seminario de la Estadística Oficial en el Ámbito Internacional: Logros y Retos Aplicables a América Latina. *Magazine Ib de la Gestión Estadística*, 5, pp. 53-61.
- Bowley, A.L. (1906). Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. *Journal of the Royal Statistical Society*, 69, pp. 548-557.
- (1913). Working-class Households in Reading. *Journal of the Royal Statistical Society*, 76, pp. 672-701.
- Brewer, K.R.W. (1963). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, 5, pp. 93-105.
- (1995). Combining design-based and model-based inference. *Business Survey Methods*, ed. by B.G. Cox, D.A. Binder, B.N. Chinappa, A. Christianson, M.J. Colledge and P.S. Kott. John Wiley and Sons, pp. 589-606.
- Cochran, W.G. (1942). Sampling theory when the sampling units are of unequal sizes. *Journal of the American Statistical Association*, 37, pp. 199-212.
- (1977). *Sampling Techniques*. 3rd edition, Wiley, New York.
- DANE (2011a). Code of good practice in statistics for Latin America and the Caribbean. Disponible en: http://www.dane.gov.co/files/noticias/BuenasPracticas_en.pdf
- (2011b). Metodología aseguramiento de la calidad de la información estadística. Versión 2.0. Disponible en: <http://unstats.un.org/unsd/dnss/docs-nqaf/Assurance%20quality%20methodology.pdf>
- Eurostat (2007). Handbook on data quality assessment methods and tools. Disponible en: <http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf>
- Fondo Monetario Internacional (2003). Marco para evaluar la calidad de las estadísticas (MECAD) de finanzas públicas. Disponible en: http://dsbb.imf.org/images/pdfs/dqrsdqaf/DQAF_GFS_Spanish.pdf
- González, M.E.; Ogus, J.L.; Shapiro, G. y Tepping B.T. (1975). Standards for the discussion and presentation of errors in surveys and census data. *Journal of the American Statistical Association*, 73, pp. 7-15.
- Groves, R.M.; Fowler, F.J.; Couper, M.P.; Lepkowski, J.M.; Singer E. and Tourangeau R. (2009). *Survey Methodology*. Wiley Series in Survey Methodology. John Wiley and Sons. Hoboken, New Jersey.
- Hansen, M.H. and Hauser, P.M. (1945). Area sampling: Some principles of sample design. *American Association for Public Opinion Research*, 9(2), pp. 183-193.
- Hansen, M.H. and Hurvitz, W.N. (1943). On the theory of sampling from a finite population. *Annals of Mathematical Statistics*, 14, pp. 333-362.
- Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, pp. 663-685.
- INEGI (2010). Norma técnica para la generación de estadística básica. Diario Oficial de la Federación. Disponible en: http://dof.gob.mx/nota_detalle.php?codigo=5167222&fecha=12/11/2010
- Jeffrey, R.C. (1992). *Probability and the Art of Judgement*. Cambridge University Press.
- Kiaer, A.N. (1895). Observations et experiences concernant des dénombrements représentatives. *Bulletin of the International Statistical Institute*, IX, Book 2, pp. 176-183.
- (1897). Sur les methodes representatives ou typologiques appliquees a la statistique. *Bulletin of the International Statistical Institute*, XI, pp. 180-189.
- (1901). Sur les methodes representatives ou typologiques. *Bulletin of the International Statistical Institute*, Book 3, pp. 66-78.
- (1997 reprint). Den representative undersokelsemetode. Christiania Videnskabselskabets Skrifte. II. Historiskfilosofiske klasse, Nr 4 (1897). English translation: The Representative Method of Statistical Surveys, Statistics Norway.
- Kruskal, W. and Mosteller, F. (1979a). Representative sampling I: Non-scientific literature. *International Statistical Review*, 47, 13-24.

- (1979b). Representative sampling II: Scientific literature, excluding statistics. *International Statistical Review*, 47, 111-127.
- (1979c). Representative sampling III: The current statistical literature. *International Statistical Review*, 47, 245-265.
- (1980). Representative sampling IV: The history of the concept in statistics, 1895-1939. *International Statistical Review*, 48, 169-195.
- Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-370.
- Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, pp. 558-606.
- Nirel R. y Glickman H. (2009). Sample surveys and censuses. In: Pfeffermann D. and Rao C.R. (eds) *Sample Surveys: Design, Methods and Applications*. Handbook of Statistics. Volume 29 A. North Holland, Elsevier, Amsterdam, The Netherlands.
- OECD (2011). Quality framework and guidelines for OECD Statistical Activities. Disponible en: <http://www.oecd.org/std/qualityframeworkforocdstatisticalactivities.htm>
- Patterson, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, Series B, 12, 241-255.
- Pfeffermann, D y Rao C.R. (2009). Sample Surveys: Design, Methods and Applications. *Handbook of Statistics*. Volume 29 A. North Holland, Elsevier, Amsterdam, The Netherlands.
- Royal, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, pp. 377-387.
- Sanz, B. (2006). Algunas reflexiones sobre el marco legal para la utilización de la información administrativa con fines estadísticos. Instituto de Estudios Fiscales. Madrid, España. Disponible en: http://www.ief.es/documentos/investigacion/seminarios/internacional/Sem_Inter_estadisticas_Sanz.pdf
- Sarndal, C.E., Swensson, B. y Wretman J. (1992). *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer Verlag. New York.
- Seng, Y.P. (1951). Historical survey of the development of sampling theories and practice. *Journal of the Royal Statistical Society*, Series A, 114, 214-231.
- Smith, T.M.F. (1978). The foundations of survey sampling: A review. *Journal of the Royal Statistical Society*, Series A, 139, 183-204.
- Statistics Canada (1987). Statistics Canada's policy on informing users of data quality and methodology. *Journal of Official Statistics*, 3, pp. 83-92.
- Statistics South Africa (2010). South African statistical quality assessment framework (SASFAQ). 2nd. Ed./Statistics South Africa-Pretoria: Statistics South Africa. Disponible en: http://www.statssa.gov.za/inside_statssa/standardisation/SASQAF_Edition_2.pdf
- UK Statistics Authority (2009). Code of practice for official statistics. Edition 1.0. Disponible en: <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/>
- United Nations (1964). Recommendations for the preparation of sample survey reports (Provisional Issue). Statistical Papers, Series C, No. 1, Rev. 2, New York: United Nations.
- United Nations Economic Commission for Europe (UNECE) (2006). Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva. Disponible en: www.unece.org/stats/documents/ece/ces/ge.41/2006/zip.1.e.pdf.
- U.S. Department of Commerce (1978a). An error profile: employment as measured by the current population survey. Statistical Policy Working Paper No. 3, Washington D.C.: U.S. Government Printing Office.
- (1978b). Glossary of nonsampling error terms: an illustration of a semantic problem in statistics (authors preface). Statistical Policy Working Paper No. 4, Washington D.C.: U.S. Government Printing Office. Disponible en: <http://ntl.bts.gov/lib/7000/7200/7203/sw4.html>

Diferencial salarial entre sindicalizados y no sindicalizados: un análisis para los sectores público y privado

José Daniel Trujillo Suárez



Diferencial salarial entre sindicalizados y no sindicalizados: un análisis para los sectores público y privado

José Daniel Trujillo Suárez: Economista de la Universidad Javeriana, magíster en economía de la Universidad de los Andes. Profesional de la Dirección de Metodología y Producción Estadística del DANE. Correo electrónico: jdtrujillos@dane.gov.co

Fecha de recepción: 17 de mayo de 2013
Fecha de aceptación: 20 de septiembre de 2013

Resumen: en este trabajo se estima el efecto que tienen los sindicatos sobre el salario de los trabajadores evaluando su distribución entre trabajadores públicos y privados. Se utiliza la metodología de Blinder-Oxaca, pero teniendo cuatro grupos diferentes: trabajadores públicos sindicalizados, trabajadores públicos no sindicalizados, trabajadores privados sindicalizados y trabajadores privados no sindicalizados. Se encuentra que la brecha salarial de sindicalización es aproximadamente 50 % menor para los trabajadores públicos que para los privados. Se observa también un diferencial positivo de 15 % del ingreso por hora para los trabajadores sindicalizados incluyendo los dos sectores.

Palabras clave: sindicatos, trabajo público y privado, brecha salarial entre sindicalizados y no sindicalizados.

Abstract: using Blinder-Oaxaca methodology this paper explores the effect of unions on wages of workers assessing their distribution between public and private workers. With four different groups (union public members, non-union public workers, union private members, non-union private workers) the results indicate that the unionization wage gap is 50 % larger in terms of wages for private sector workers and there is also a 15 % positive differential of income per hour for union workers including the two sectors.

Key words: unions, public and private work, wage gap between union and non-union workers.

1. Introducción

Dentro de la economía laboral se le ha dado particular importancia al efecto que tienen los sindicatos sobre la remuneración a sus afiliados. En diversos trabajos se muestra el diferencial positivo que tienen estos trabajadores sobre los que no están sindicalizados, diferencial también llamado prima sindical o de sindicalización. Mankiw (2000) presenta los sindicatos como una de las principales causas de las rigideces salariales (junto al salario mínimo y los salarios de eficiencia) que son a su vez las causantes del desempleo en espera, en donde el salario observado está consistentemente por encima del salario de equilibrio que iguala la oferta a la demanda de empleo. Los sindicatos logran mayores beneficios eliminando la competencia de los desempleados¹ por medio del control de la oferta de trabajo, la acción política para limitar la oferta de trabajo y con las huelgas (Cúellar, 2009). Esta última afecta el salario a través de la reducción de la mano de obra o mediante la negociación sobre el salario directamente; aunque el éxito para controlar la oferta de trabajo depende de la elasticidad de demanda de su producto, la elasticidad de sustitución trabajo-capital, el grado de competencia de la industria, las tasas de interés y el salario que está dispuesto a pagar el empleador. Para comprobar esto se ha buscado evidencia empírica para probar que la prima sindical es positiva. Dado que las condiciones legales de los sindicatos y la estructura laboral no son iguales en todos los países, es importante hacer diferentes estimaciones con distintas metodologías que permitan verificar si en realidad los sindicatos están siendo parte de las rigideces

salariales, las cuales mantienen la tasa de desempleo elevada.

Los niveles de sindicalización se han mantenido bajos en Colombia, consecuencia —según los sindicalistas— de la cultura antisindical que predomina en los sectores empresariales y en el estado, que se han resistido al derecho de asociación como un derecho fundamental (Vázquez, 1999). No obstante, las conclusiones de la literatura sobre el diferencial salarial que tienen los trabajadores sindicalizados muestran efectos positivos y significativos, por lo que la elección racional de los trabajadores debería ser la afiliación sindical. Otra razón para que los niveles de sindicalización se mantengan bajos es que los empleadores ofrecen beneficios a los trabajadores para disuadirlos de sindicalizarse; por lo que la brecha salarial de sindicalización sería menor², y a su vez, con menos incentivos para la sindicalización. Los empleadores están dispuestos a pagar cierta prima sindical siempre que recuperen el poder de negociación y eviten amenazas de huelgas que conlleven a una disminución en la producción y en los beneficios.

La distribución de trabajadores públicos y privados dentro de los sindicalizados no es la misma que dentro de los no sindicalizados, por esto es importante tener en cuenta que la brecha sindical puede estar influenciada por la composición de los sectores en cada uno de estos grupos. De acuerdo con la Gran Encuesta Integrada de Hogares, entre 2008 y 2011 el 26 % de los trabajadores públicos estaban sindicalizados, mientras que para los privados este porcentaje era tan sólo del 1,5.

Debido a las diferentes composiciones en la sindicalización en la

parte pública y en la privada, es importante conocer qué está pasando con esta brecha salarial dentro de cada uno de los grupos. Por ello, en este trabajo se buscará ver qué tan favorable es para los empleados públicos la diferencia salarial entre sindicalizados y no sindicalizados en comparación con los trabajadores del sector privado; adicionalmente se utiliza la descomposición de Nopo (2008) para calcular la prima sindical. Las metodologías aplicadas aquí no se habían utilizado antes para realizar el cálculo planteado.

Se utilizará Blinder-Oxaca escogiendo dos grupos de tratamiento (normalmente es con un grupo): estatus sindical y sector; de manera que se tenga en cuenta la importancia de las características observadas de los trabajadores en la posibilidad de pertenecer a cada uno de los grupos o viceversa. De esta manera, la metodología propuesta presenta beneficios adicionales ya que los estimadores de las variables de interés no incluyen a las diferencias de los estimadores de las características observadas de éstos grupos, evitando así sesgos subyacentes. En todos los años se encuentra que la brecha salarial de sindicalización³ para los trabajadores públicos es menor que para los privados, en promedio esta diferencia es cerca del 50 % cuando se mira el ingreso mensual, y alrededor de 30 % con el ingreso por hora; además los trabajadores sindicalizados ganan en promedio 15 % más que los no sindicalizados.

Inicialmente los trabajos que estudiaban la brecha sindical suponían que el estatus sindical era una variable exógena. Estudios posteriores señalaron esta forma de inclusión como errónea ya que estar o no sindicalizado es una decisión tomada por el individuo y a su vez depende del ingreso esperado (sesgo de selección) y la correlación en-

¹ Los desempleados presionan los salarios a la baja debido al exceso de oferta de trabajo.

² Si todos los empleadores actuaran de esta manera, la diferencia entre ingresos de sindicalizados y no sindicalizados sería nula, eliminando así los incentivos para afiliarse a un sindicato.

³ Se llamará brecha salarial de sindicalización o brecha sindical a la diferencia en la remuneración de los trabajadores sindicalizados y no sindicalizados.

tre el estatus sindical y el término de error (Abowd and Farber, 1982 y Freeman, 1983). Sin embargo, otros trabajos suponen que el efecto de variables no observables sobre la afiliación sindical es reducido, teniendo en cuenta que el objetivo primario de los individuos no es encontrar un empleo con sindicato sino que ellos se afilian de acuerdo con la representación sindical dentro de la empresa (Cuesta, 2005). Inicialmente se utilizará el supuesto de que el status sindical es una decisión exógena, ya que esta depende en mayor medida del tipo y tamaño de empresa en donde pueden o no haber sindicatos establecidos. Después se eliminarán a las personas que tengan menos de tres años trabajando en esta empresa para evaluar la robustez de los resultados reduciendo la endogeneidad en la decisión de sindicalizarse y el ingreso actual.

Este trabajo está compuesto por cinco secciones. La primera es esta introducción. En la segunda se realiza una revisión de trabajos que estudian la brecha sindical en diferentes países y en Colombia, así como también los determinantes del estatus sindical. En la tercera se explica la metodología y se muestra una descripción de los datos que se utilizarán. En la cuarta se presentan los resultados principales, y en la quinta se concluye.

2. Revisión de literatura

2.1. Efecto de estar sindicalizado sobre el ingreso

Robinson (1984) estudió el diferencial salarial entre los sectores pú-

blico y privado teniendo en cuenta si el trabajador es o no sindicalizado. Este trabajo se realizó para Canadá con datos de la Social Change in Canadian Survey para 1979. El autor encontró evidencia de mayores salarios para los sindicalizados en el sector público; sin embargo, cuando se controla por sindicatos se reduce el diferencial salarial entre los sectores.

Gyourko and Tracy (1986) realizaron un estudio sobre la determinación salarial entre los sectores privado/público y los trabajadores sindicalizados/no sindicalizados. Los autores reconocen la endogeneidad de las decisiones de pertenecer a cada uno de estos grupos —a diferencia de lo sucedido con la literatura anterior a esta—, por lo que realizan un logit multinomial encontrando sesgo de selección en los trabajadores privados no sindicalizados y los públicos sindicalizados. En este trabajo se presenta evidencia de un diferencial importante para los trabajadores públicos sindicalizados.

En un estudio realizado en Australia (uno de los países con mayor densidad sindical), Christie (1992) encuentra que el diferencial salarial agregado entre sindicalizados y no sindicalizados fue de 15,3 %, resaltando a su vez la importancia relativa de las variables de empresa en comparación con las individuales para la determinación de la probabilidad de pertenecer o no a un sindicato.

Kuhn (1998) encuentra que en promedio los sindicatos hacen que los salarios de los trabajadores canadienses y estadounidenses sean un 15 % más altos y que estos a su vez presentan una menor dispersión, reduciendo la importancia del desempeño y las habilidades individuales. Adicionalmente el estudio indica que los sindicatos no tienen efecto negativo sobre el empleo ni sobre el salario de los no sindicalizados.

Chrysanthou (2008), con datos de la British Household Panel Survey entre 1991 y 2003, utilizó un modelo dinámico de estatus sindical y determinación salarial para estimar el impacto de los sindicatos en Gran Bretaña. El autor encontró que las variables no observadas que afectan a la afiliación sindical también afectan la magnitud del impacto de los sindicatos en los salarios. Además mostraron que en un primer período de estudio (1991-1996) el efecto de los sindicalizados hombres era 9,2 % mientras que entre 1997 y 2003 este efecto era tan sólo de 5,8 %; de igual forma para las mujeres este cambio fue de 17,6 % a 11,5 %. Calcularon también estimadores del efecto sindical por medio de efectos fijos y encontraron que éste era mucho más bajo, por lo que este tipo de metodología sirve para calcular el límite inferior del estimador del efecto de los sindicatos sobre los salarios (Freeman, 1983).

Landerretche et al. (2010), utilizando un panel de datos para Chile, estiman el premio en términos de ingresos generados por el sindicato por medio de una corrección de dos etapas para estimar consistentemente el efecto de la variable categórica de sindicato, y concluyen que dicho premio es del 4,5 %. La primera etapa consiste en una estimación de un probit con efectos aleatorios en donde se determina la posibilidad de estar sindicalizado, y en la segunda etapa se utiliza esta estimación o corrección de la primera para reemplazar la variable endógena —estar sindicalizado— en la determinación del salario.

Cuesta (2005) estima el efecto diferencial de los salarios para las personas sindicalizadas en Colombia utilizando la Encuesta Nacional de Hogares para el año 2000. Concluye que los salarios de los sindicalizados, manteniendo todos los otros efectos constantes y usando *propensity score matching*⁴,

⁴Es un método de evaluación de impacto que consiste en encontrar un grupo de control comparable con el grupo de tratamiento basándose en características observables de forma que al evaluar el programa las muestras sean comparables.

son entre 3 % y 5 % más altos que los de los no sindicalizados. Este efecto es aún más pronunciado en el sector privado, donde el salario de los primeros es en promedio 25 % más alto que el de los no sindicalizados. Cuando estima el modelo para trabajadores privados y públicos, normalmente los resultados no son significativos, concluyendo que el hecho de tener trabajadores del gobierno disminuye la brecha de salarios entre sindicalizados y no sindicalizados, ya que éstos normalmente se rigen por escalas salariales ya predeterminadas.

Trujillo (2011), utilizando información de las convenciones colectivas firmadas entre 2006 y 2009, muestra que el efecto de los sindicatos sobre los salarios no solo se mira como una variable discreta, sino que empresas que tienen una mayor proporción de empleados sindicalizados tienden a firmar convenciones colectivas con un mayor crecimiento salarial. Además, se muestra que cuando hay desaceleración económica esta diferencia salarial entre las empresas de mayor y menor proporción de sindicalizados se vuelve más amplia.

Todos los trabajos anteriores, cuyas metodologías de estimación son diferentes, muestran consenso del efecto de estar sindicalizado sobre los salarios, sin importar el sector a que pertenezcan, el sexo u otras variables económicas y sociodemográficas. Dado este resultado, las bajas tasas de sindicalización en Colombia se relacionan con barreras de entrada que ponen en entredicho la endogeneidad del estatus sindical de los trabajadores.

2.2. Determinantes de sindicalización

Dado que no hay consenso sobre la endogeneidad del estatus sindical, es importante conocer diferentes estudios que se han enfocado en encontrar la importancia y el

efecto de diferentes determinantes sobre estar sindicalizado o no. Por ejemplo, Schur and Kruse (1992) estudian las razones por las cuales las mujeres son menos propensas a estar sindicalizadas que los hombres, descartando el hecho de que menores tasas de sindicalización sean producto del menor deseo de estar sindicalizadas. De hecho, los autores encontraron que el grupo de trabajadores mujeres no sindicalizadas reportó tener una buena percepción de los sindicatos y creen que éstos pueden ayudar a mejorar las condiciones salariales, mientras los hombres no sindicalizados mostraron poco interés en los sindicatos. Por esto la menor tasa de sindicalización de mujeres en el sector privado recae sobre barreras inherentes al sexo del trabajador mientras la decisión de estar sindicalizado para los hombres está casi perfectamente correlacionada con la tasa de sindicalización de estos.

Igualmente para Colombia, Guataquí et al. (2009) encuentran que los hombres tienen mayor probabilidad de estar sindicalizados que las mujeres. Afectan también positivamente esta probabilidad una mayor escolaridad y experiencia, ser empleado público en empresas grandes y con contratos a término indefinido.

Así mismo hay trabajos similares a los dos anteriores que muestran la importancia de las variables socioeconómicas sobre la probabilidad de estar sindicalizado; sin embargo, como se mencionaba anteriormente, es importante tener en cuenta la facilidad con la que los trabajadores pueden llevar a cabo el deseo de sindicalización con barreras de acceso tan pronunciadas en un país como Colombia (Vázquez, 1999).

2.3. Convenciones colectivas

Las convenciones colectivas son contratos que se firman entre un número determinado de empleados (sindicatos) y empleadores.

Existen diferentes tipos de sindicatos; entre ellos los de empresa, industria, gremiales, federaciones y confederaciones. Los empleados deciden a qué sindicato pertenecer, siendo los más comunes los sindicatos a nivel de empresa. Si las personas no quieren pertenecer a estos sindicatos, entonces pueden inscribirse a los de industria y acceder a los beneficios que posiblemente adquieran de una negociación colectiva con mayor poder de negociación, siempre y cuando la empresa sea del tamaño suficiente y tenga la capacidad técnica para poder cubrir estos costos.

Dentro del contrato se especifica la cantidad del aumento salarial para cada año, al igual que las prestaciones legales y extralegales. La ley da independencia a estos contratos mientras no se viole lo estipulado con las personas que ganan el salario mínimo.

Hay algunas excepciones en las cuales las personas sin pertenecer a los sindicatos pueden hacer parte de los beneficios de la convención; por ejemplo, si al momento de firmar con uno o más sindicatos, y este o estos están conformados por más de la tercera parte de las personas que trabajan en esta o estas empresas, los beneficios y obligaciones se extienden al resto de trabajadores.

Además, las convenciones que benefician a algunas personas (que no están afiliadas a un sindicato) no se aplican cuando previamente estén recibiendo beneficios superiores. Si hay una convención que está conformada por más de dos terceras partes de los trabajadores de esa rama industrial, entonces esta convención rige para todas las demás empresas siempre y cuando estas tengan la capacidad técnica y económica y los trabajadores que inicialmente no están en la convención no tengan mejores condiciones. El gobierno es el

que decide la capacidad técnica de la empresa en cada rama industrial debido a los incentivos que tengan algunas empresas de catalogarse con la accesibilidad de cumplir o no con las convenciones colectivas de industria.

Este tipo de contratos se siguen cumpliendo sin importar que el sindicato de empleados o patronal que firmó la convención colectiva se disuelva. Aunque la duración de los convenios normalmente es de dos años, puede haber casos en que esta no se estipule en el documento; entonces se alargaría de seis meses en seis meses hasta que se firme una nueva convención. Esta misma prórroga se da si antes de sesenta días de terminada, alguna de las partes no manifiesta la terminación del contrato⁵.

2.4. Pactos colectivos

Son pactos entre empleadores y trabajadores no sindicalizados que se pueden celebrar cuando el sindicato de empresa no agrupa más de una tercera parte de los trabajadores. En Colombia se ha utilizado esta figura para evitar que los trabajadores se sindicalicen y puedan ejercer el derecho a la huelga con el fin de persuadir al empleador para aumentar los salarios o beneficios de los trabajadores.

3. Metodología y datos

¿Los empleados públicos tienen mayores beneficios que los privados por estar sindicalizados? Para responder a esta pregunta se utilizará un Blinder-Oxaca con dos grupos, como se describe en Rodríguez-Lesmes et al. (2011), donde estos grupos son el estatus sindical y el sector donde se desempeñan. La siguiente ecuación muestra lo que se estimará:

$$y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 S_i + \alpha_3 T_i S_i + \alpha_4 X_i + \alpha_5 X_i T_i + \alpha_6 X_i S_i + \alpha_7 X_i T_i S_i + \varrho_i \quad (1)$$

T_i es uno si el empleado es público y cero en caso contrario; Si es uno si el trabajador es sindicalizado y cero si no lo es X_i es un vector de características de la persona i , siendo α_4 un vector a diferencia de los otros coeficientes; y, y_i es el logaritmo del salario por hora.

Además,

$$\varrho_i = \eta_i + \mu_s + \omega_t$$

$$E(\varrho_i) = \mu_s + \omega_t$$

Como algunas variables del vector X están correlacionadas con los tratamientos, sindicalizado o trabajador público, entonces $E(X|T, S) \neq E(X)$. De esta forma, como notación se tomará al $E(X|T = t, S = s) = X_{ts}$ con $st = 0, 1$.

El valor esperado de la ecuación 1 dado cada uno de los tratamientos es (excluyendo subíndice i por simplicidad):

$$E(Y|T = 0, S = 0) = \alpha_0 + \alpha_4 X_{00}$$

$$E(Y|T = 1, S = 0) = \alpha_0 + \alpha_1 + \alpha_4 X_{10} + \alpha_5 X_{10} + \omega_t$$

$$E(Y|T = 0, S = 1) = \alpha_0 + \alpha_2 + \alpha_4 X_{01} + \alpha_6 X_{01} + \mu_t \quad (2)$$

$$E(Y|T = 1, S = 1) = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 X_{11} + \alpha_5 X_{11} + \alpha_6 X_{11} + \alpha_7 X_{11} + \omega_t + \mu_t$$

De la ecuación 2 se pueden extraer dos diferencias importantes; la diferencia entre sindicalizados y no sindicalizados para trabajadores privados:

$$E(Y|T = 0, S = 1) - E(Y|T = 0, S = 0) = (\alpha_0 + \alpha_2 + \alpha_4 X_{01} + \alpha_6 X_{01} + \mu_t) - (\alpha_0 + \alpha_4 X_{00})$$

y la diferencia entre sindicalizados y no sindicalizados para trabajadores públicos:

$$E(Y|T = 1, S = 1) - E(Y|T = 1, S = 0) = (\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 X_{11} + \alpha_5 X_{11} + \alpha_6 X_{11} + \alpha_7 X_{11} + \omega_t + \mu_t) - (\alpha_0 + \alpha_1 + \alpha_4 X_{10} + \alpha_5 X_{10} + \omega_t)$$

Si los trabajadores públicos tienen una influencia adicional sobre los privados en cuanto a la brecha de sindicalización, entonces la diferencia entre las dos expresiones anteriores debe resultar en un valor positivo (se querrá mostrar que es

significativo). Por otro lado, si los trabajadores privados son los que tienen influencia adicional en la brecha de ingresos de los sindicalizados y no sindicalizados, la diferencia será negativa. La diferencia total se define como:

$$\delta = [E(Y_i|T = 1, S = 1) - E(Y_i|T = 1, S = 0)] - [E(Y_i|T = 0, S = 1) - E(Y_i|T = 0, S = 0)]$$

Y de acuerdo con las ecuaciones 1 y 2, δ se puede expresar como:

$$\delta = \alpha_4 [(X_{11} - X_{10}) - (X_{01} - X_{00})] + \alpha_3 + (\alpha_5 + \alpha_6 + \alpha_7) X_{11} - \alpha_5 X_{10} - \alpha_6 X_{01} \quad (3)$$

Acá se puede observar que δ no contiene a μ_s y ω_t , expresiones del término error inherentes a cada grupo que pueden afectar los estimadores de interés en un OLS. Si no existiera correlación entre los tratamientos y algunas variables explicativas entonces $E(X|T = t, S = s) = X$ y δ sería igual a α_3 , similar al estimador de diferencias en donde los efectos fijos a nivel de grupo se eliminan, como sucede acá. Por otro lado, si se tuviera un sólo grupo de interés (por ejemplo, público y privado), δ sería $E(Y_i|T = 1) - E(Y_i|T = 0)$ y la expresión quedaría de la siguiente forma:

$$\delta = \alpha_1 + \alpha_5 X_1 + \alpha_4 (X_1 - X_0)$$

Esta es la conocida expresión de Blinder-Oxaca en donde esta diferencia se descompone en diferencia explicada ($\alpha_4 (X_1 - X_0)$) y diferencia no explicada ($\alpha_1 + \alpha_5 X_1$).

La ecuación 3 es una expresión similar a un Blinder-Oxaca pero incluyendo los dos grupos mencionados en vez de uno. De igual manera δ se puede descomponer en diferencia explicada:

$$\Delta X = \alpha_4 [(X_{11} - X_{10}) - (X_{01} - X_{00})]$$

y diferencia no explicada:

$$\Delta O = \alpha_3 + (\alpha_5 + \alpha_6 + \alpha_7) X_{11} - \alpha_5 X_{10} - \alpha_6 X_{01}$$

⁵ Código sustantivo del trabajo, Título III, Capítulo I, artículos 467, 470, 471, 472 y 478.

Δx puede ser interpretada como la diferencia en las características promedio de los dos grupos mientras $\Delta 0$ refleja la diferencia que no puede ser explicada siendo un efecto inherente al grupo de trabajadores públicos sindicalizados.

En la sección 4.2 se utilizará un emparejamiento de tal forma que se pueda dividir de manera más adecuada las diferencias explicadas y no explicadas sin incluir observaciones que tengan diferentes características observadas. La variable de tratamiento será el estatus sindical y se realizarán estimaciones para el sector público y privado así como también para otras características de los trabajadores. La idea de esto es proveer una metodología alternativa para ver la robustez de los resultados; sin embargo, cabe resaltar la imposibilidad de corroborar la significancia de estos cuando se miran por grupos (como público y privado), por lo que esto sólo servirá para reafirmar la dirección de los efectos.

Datos

La base de datos de este trabajo se construyó a partir de la Gran Encuesta Integrada de hogares del DANE entre los años 2008 y 2011.

Para el ejercicio estadístico se realizan algunos filtros con el objetivo de evitar heterogeneidades y *outlayers*. En la muestra final se dejan los trabajadores particulares, del gobierno y cuenta propia debido a diferencias importantes en el tipo de empleo de los jornaleros, trabajadores domésticos, patrones y trabajadores sin remuneración. Se utilizan los ingresos calculados por el DANE con sus imputaciones respectivas incluyendo solo el ingreso por primera actividad para toda la muestra y el ingreso en especie para el caso de los trabajadores particulares y del gobierno. Se eliminan 213 observaciones que no reportan nivel educativo, los trabajadores menores a 19 años y mayores a 65, y 179 observaciones que no tenían información clara del sector económico, así como 10.430 que reportaban ingreso 0 (debido a que se utilizará el logaritmo de este).

En el cuadro 1 se muestra la distribución de los trabajadores sindicalizados y no sindicalizados en los sectores público y privado.

De los sindicalizados, la mayoría se encuentran en el sector público (56 %); sin embargo, sólo 6 % de los empleados trabajan en este sector y tan sólo 2,7 % de los trabajadores estudiados se encuentran sindicalizados.

(sindicalizados, no sindicalizados, trabajadores públicos y privados) por sexo, nivel educativo, rangos de edad, meses que lleva trabajando y horas trabajadas a la semana.

Como es de esperarse, a mayor educación mayor el ingreso por hora para todos los grupos; resaltándose la importancia marginal de pasar de bachillerato a educación superior. En cuanto al sexo del trabajador, las diferencias no parecen ser significativas. Incluso para el grupo de trabajadores del sector público y sindicalizados, los salarios por hora en promedio son exactamente los mismos, mientras que para los públicos no sindicalizados las mujeres estarían más favorecidas.

En todos los grupos los más jóvenes son los que ganan menos, y su diferencia con el siguiente rango de edad es más llamativa para los trabajadores públicos que no están sindicalizados; así mismo, más experiencia en este empleo significa más ingreso, salvo en algunos casos como el de los privados no sindicalizados en donde los que llevan más de dos años ganan en promedio menos que los que llevan entre uno y dos años. Parece haber una relación monótonica decreciente entre el ingreso por hora y las horas trabajadas a la semana ya que las personas que más ganan son las que menos trabajan; sin embargo, éste patrón se marca por las grandes diferencias entre los trabajadores públicos, ya que para los privados los que trabajan menos de 25 horas ganan ligeramente menos que los que trabajan entre 25 y 48 horas.

Finalmente, en todos los grupos se ha visto un incremento de la remuneración a través del tiempo. En promedio para toda la muestra, los trabajadores públicos sindicalizados son los que más ganan, seguidos por los públicos no sindicalizados. Los que menos ganan son los privados no sindicalizados.

Cuadro 1. Distribución de estatus sindical y sector. 2008-2011

		Sector privado	Sector público	Total
		(%)	(%)	(%)
Sindicalizado	Sí	56.0	44.0	100.0
	No	4.5	95.5	100.0
	Total	5.9	94.1	100.0
Sindicalizado				
Sector		Sí	No	Total
	Público	26.0	74.0	100.0
	Privado	1.3	98.7	100.0
	Total	2.7	97.3	100.0

Fuente: GEIH 2008-2011 y cálculos del autor.

En la tabla 1 se muestran las estadísticas descriptivas de las principales variables utilizadas.

En la tabla 2 se presentan los promedios del ingreso por hora de cada uno de los grupos utilizados

Tabla 1. Variables descriptivas muestra final.

Variable	Obs	Media	Desviación	Min	Max std.
Mujer	1073826	0.43	0.4952	0.00	1.00
Edad	1073826	38.20	11.7056	19.00	65.00
Edad cuadrado	1073826	1596.06	943.4646	361.00	4225.00
Primaria	1073826	0.30	0.4566	0.00	1.00
Secundaria	1073826	0.33	0.4709	0.00	1.00
Superior	1073826	0.23	0.4234	0.00	1.00
Casado	1073826	0.26	0.4401	0.00	1.00
Cabeza de hogar	1073826	0.50	0.5000	0.00	1.00
Horas semanales	1073826	47.76	17.8783	1.00	130.00
Tiempo trabajando	1073826	80.98	102.1421	0.00	660.00
Núm. trabajadores empresa	1073826	3.96	3.3312	1.00	9.00
Asalariados familia	1073826	0.25	0.2626	0.00	1.00
Cuenta propia familia	1073826	0.27	0.2594	0.00	1.00
Desempleados familia	1073826	0.04	0.1040	0.00	0.86
Menores familia	1073826	0.07	0.1255	0.00	0.80
Mayores familia	1073826	0.04	0.1052	0.00	0.80
Minería	1073826	0.01	0.0910	0.00	1.00
Industria liviana	1073826	0.07	0.2625	0.00	1.00
Industria pesada	1073826	0.05	0.2241	0.00	1.00
Servicios públicos	1073826	0.01	0.0986	0.00	1.00
Construcción	1073826	0.06	0.2450	0.00	1.00
Comercio	1073826	0.30	0.4587	0.00	1.00
Transporte	1073826	0.11	0.3175	0.00	1.00
Sector financiero	1073826	0.09	0.2858	0.00	1.00
Servicios sociales	1073826	0.23	0.4181	0.00	1.00
Log. ingreso mensual	1073826	13.10	0.9887	8.11	18.32
Log. ingreso por hora	1073826	7.94	0.9265	2.79	13.30

Fuente: GEIH.

En el cuadro 2 se muestran los porcentajes de los sindicalizados y no sindicalizados por sexo, grupos de edad y educación.

Las personas que tienen mayor educación son más propensas a estar sindicalizadas; 65 % de los que están sindicalizados tienen educación superior, mientras que en la muestra este porcentaje es solo de 21 %. Así mismo, en la muestra los que no tienen educación alguna son 17 % y en el grupo de sindicalizados estos tan solo representan el 6 %.

Contrario a la literatura, los hombres representan una menor proporción dentro de los sindicaliza-

dos que dentro de la muestra total y dentro del grupo de edades. Los que tienen mayor propensión a sindicalizarse son las personas mayores a 35 años, sobre todo los trabajadores entre 46 y 55 años.

4. Resultados

4.1. Diferencia de la brecha sindical entre público y privado

En la tabla 3 se muestran los resultados de la diferencia total, explicada y no explicada⁶ utilizando la metodología descrita en la sección anterior y con variable dependiente el ingreso mensual.

Para todos los años la diferencia total (δ) es negativa, así como sus componentes. Se puede observar también que esta diferencia tiene magnitud cada vez mayor a medida que pasan los años. En todos los períodos casi toda la diferencia total viene de la diferencia explicada por las variables observables mientras que una pequeña parte se le atribuye a factores no observados y al hecho inherente de ser un trabajador público sindicalizado. En general, la brecha de sindicalización es aproximadamente 50 % menor para los trabajadores públicos.

En la tabla 4 se muestran resultados similares a la tabla anterior pero esta vez la variable dependiente es el logaritmo del salario por hora.

⁶ Estas diferencias se interpretan como el cambio porcentual en el ingreso respectivo; por ejemplo, una diferencia de 0,1 representa un aumento del 10 %.

Tabla 2. Ingreso-hora por sector y estatus sindical.

	Sindicalizado Público	No sindicalizado Público	Categoría Sind. Privado	No sind. Privado	Total
Educación					
Ninguna	8.40	8.06	7.22	7.23	7.24
Primaria	8.70	8.30	7.83	7.59	7.60
Secundaria	8.93	8.55	8.29	7.92	7.96
Superior	9.45	9.23	9.27	8.70	8.83
Total	9.35	8.96	8.39	7.82	7.90
Sexo					
Mujer	9.35	9.03	8.44	7.77	7.88
Hombre	9.34	8.90	8.37	7.86	7.92
Total	9.35	8.96	8.39	7.82	7.90
Edad					
entre 19 y 24	8.88	8.23	8.00	7.70	7.71
entre 25 y 35	9.08	8.82	8.33	7.94	7.99
entre 36 y 45	9.29	9.01	8.43	7.87	7.96
entre 46 y 55	9.41	9.16	8.49	7.78	7.92
más de 55	9.56	9.31	8.34	7.56	7.69
Total	9.35	8.96	8.39	7.82	7.90
Meses en este trabajo					
Menos de 7 meses	8.99	8.58	7.96	7.68	7.70
7 a 12 meses	9.15	8.65	8.38	7.83	7.86
12 a 24 meses	9.26	8.88	8.37	7.93	7.96
Más de 24 meses	9.36	9.04	8.45	7.86	7.98
Total	9.35	8.96	8.39	7.82	7.90
Horas trabajadas a la semana					
Menos de 25	9.83	9.30	8.44	7.82	7.84
24 a 28	9.37	9.08	8.54	7.95	8.06
Más de 48	8.97	8.59	8.09	7.63	7.67
Total	9.35	8.96	8.39	7.82	7.90
Año					
2008	9.19	8.84	8.29	7.76	7.84
2009	9.33	8.93	8.26	7.78	7.86
2010	9.40	9.03	8.43	7.84	7.91
2011	9.48	9.08	8.54	7.90	7.98
Total	9.35	8.96	8.39	7.82	7.90

Fuente: GEIH 2008-2011 y cálculos del autor.

Nota: El ingreso-hora se define como logaritmo del ingreso mensual de la primera actividad dividido por el número de horas trabajadas al mes.

Ahora, la diferencia no explicada tiene una mayor participación que en el caso anterior. Casi una tercera parte de la diferencia total está compuesta por la diferencia no explicada. En promedio, los trabajadores públicos tienen un 30 % de menor brecha cuando el salario es por hora. Los errores de las estimaciones mostradas en las tablas 3 y 4 se calculan con *bootstrapping*.

En las tablas 5 y 6 se muestran las estimaciones de la ecuación 1,

sin incluir las interacciones de los tratamientos y las variables explicativas, calculadas por MCO. En la primera se puede observar que en promedio los trabajadores públicos (α_1) y sindicalizados (α_2) individualmente ganan más que su contraparte mientras que la interacción (α_3) entre estos trabajadores sin incluir el efecto con las variables explicativas es negativa para todos los años, aunque no significativa.

4.2. Descomposición y emparejamiento

Para estimar acertadamente el efecto que tienen los diferentes tipos de empleo sobre el ingreso de los trabajadores se utilizará adicionalmente una técnica cuasi experimental expuesta en Nopo (2008). Esta consiste en seleccionar dos grupos (por ejemplo empleados sindicalizados y no sindicalizados) eligiendo uno de tratamiento (sindicalizados); buscando para cada

Cuadro 2. Distribución de sindicalizados y no sindicalizados por sexo, edad y educación.

	Sindicalizado		
	Sí	No	Total
Educación			
Ninguna	6.0	17.4	17.1
Primaria	12.9	31.6	31.1
Secundaria	18.6	31.4	31.1
Superior	62.5	19.6	20.7
Total	100.0	100.0	100.0
Sexo			
Mujer	44.4	40.0	40.1
Hombre	55.6	60.0	59.9
Total	100.0	100.0	100.0
Edad			
entre 19 y 24	2.6	14.4	14.1
entre 25 y 35	18.3	32.0	31.6
entre 36 y 45	31.1	26.0	26.1
entre 46 y 55	34.6	19.1	19.5
más de 55	13.5	8.6	8.7
Total	100.0	100.0	100.0

Fuente: GEIH 2008-2011 y cálculos del autor.

Tabla 3. Diferencias del logaritmo del ingreso mensual.

Diferencia	2008	2009	2010	2011
Dif. Total	-0,4650***	-0,4573***	-0,5156***	-0,5322***
	(0,0000)	(0,0003)	(0,0006)	(0,0002)
$\Delta 0$	-0,0640***	-0,0439***	-0,0489***	-0,0242***
	(0,0000)	(0,0003)	(0,0006)	(0,0002)
ΔX	-0,4010***	-0,4134***	-0,4667***	-0,5080***
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Obs 281769				

Fuente: elaboración propia.

Errores estándar con Bootstrapping con 100 repeticiones. Significancia: * 10 %, ** 5 %, *** 1 %.

Tabla 4. Diferencias del logaritmo del ingreso por hora.

Diferencia	2008	2009	2010	2011
Dif. Total	-0,3101***	-0,2743***	-0,2919***	-0,2816***
	(0,0000)	(0,0003)	(0,0006)	(0,0002)
$\Delta 0$	-0,1048***	-0,0756***	-0,0937***	-0,0744***
	(0,0000)	(0,0003)	(0,0006)	(0,0002)
ΔX	-0,2053***	-0,1987***	-0,1981***	-0,2072***
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Obs 281769				

Fuente: elaboración propia.

Errores estándar con Bootstrapping con 100 repeticiones. Significancia: * 10 %, ** 5 %, *** 1 %.

una de estas observaciones todos los controles (no sindicalizados) que tengan las mismas características como edad, educación, sexo, sector económico, estado civil, posición dentro de la familia (ser

cabeza de hogar o no) y duración en este trabajo y horas trabajadas. A partir de esto se crea un individuo sintético, el cual se empareja con la correspondiente observación de tratamiento hasta que se agoten.

A partir de la anterior metodología, la brecha del ingreso entre grupos se puede explicar por cuatro factores: una brecha debida a que los grupos tienden a tener diferentes distribuciones de las características en un soporte común de características individuales; una brecha para la combinación de características que solo están en el grupo de controles; una brecha para la combinación de características que sólo están en el grupo de tratados; y una última brecha no explicada. El efecto diferencial y el de las variables no observadas (correlacionadas con los grupos) sobre el ingreso están recogidos por esta última.

En la tabla 7 de la diferencia total entre grupos y la descomposición mencionada (donde d.e es la desviación estándar del estimador de diferencia no explicada); adicionalmente los resultados se muestran por tipo de empleo, sexo, educación y rango de edad. La variable dependiente es el ingreso-hora.

En la tabla, P. trat es el porcentaje de observaciones de tratamiento que están en el soporte común y P. control es el porcentaje de observaciones de control que están en el soporte común definido por las variables mencionadas.

La brecha total está definida como $Y_t/Y_c - 1$, donde Y_t es el ingreso (o su logaritmo) de los tratados y Y_c el de los controles. En la tabla 9, se ve que la brecha total es positiva; así mismo, la parte no explicada de esta también es significativa, por lo que las condiciones laborales medidas a través del ingreso siempre son superiores para los trabajadores sindicalizados. De igual forma, esta diferencia es menor para los trabajadores públicos que para los privados; sin embargo, nuevamente la metodología tal como está no permite medir esa significancia como se hizo al principio de esta sección. Los resultados son similares a los mostrados en la sección 4.1, ya que

Tabla 5. MCO Ingreso mensual.

Variables	(1)	(2)	(3)	(4)
	2008 ingla	2009 ingla	2010 ingla	2011 ingla
Alfa1	1.014** (0.400)	0.639 (0.683)	1.459** (0.690)	1.244 (3,616)
Alfa2	0.0784 (0.208)	0.00700 (0.217)	0.0958 (0.220)	0.314 (0.197)
Alfa3		0.525	-0.265	-0.0232
Constante	10.57*** (0.0179)	(0.804) 10.55*** (0.0178)	(0.671) 10.52*** (0.0173)	(1.036) 10.58*** (0.0167)
Observaciones	260,041	260,730	271,286	281,769
R-cuadrado	0.521	0.538	0.546	0.543

Fuente: GEIH 2008-2011 y cálculos del autor.
Errores estándar en paréntesis
*** p<0.01, ** p<0.05, * p<0.1

Tabla 6. MCO Ingreso por hora.

Variables	(1)	(2)	(3)	(4)
	2008 inglahora	2009 inglahora	2010 inglahora	2011 inglahora
Alfa1	0.846** (0.392)	0.567 (0.667)	1.049 (0.675)	1.397 (3,538)
Alfa2	-0.0350 (0.204)	-0.120 (0.212)	0.0531 (0.215)	0.172 (0.193)
Alfa3		0.850	0.00389	0.440
Constante	6.635*** (0.0176)	(0.785) 6.633*** (0.0174)	(0.657) 6.625*** (0.0170)	(1.014) 6.704*** (0.0163)
Observaciones	260,041	260,730	271,286	281,769
R-cuadrado	0.493	0.502	0.495	0.487

Fuente: GEIH 2008-2011 y cálculos del autor.
Errores estándar en paréntesis.
*** p<0.01, ** p<0.05, * p<0.1

la brecha de sindicalización de los trabajadores públicos es cerca de la mitad de la brecha de los privados (0,0424 contra 0,0857).

Así mismo, en la tabla 9 se pueden observar algunos resultados por grupos. La brecha sindical es mayor para las mujeres, para los que tienen mayor educación, y para los trabajadores de mayor edad.

En el caso del ingreso por hora, los trabajadores públicos ganan en promedio más; aunque la signi-

ficancia no es estable a través de los años. La estimación a través de MCO es un primer paso para el cálculo de coeficientes debido a que bajo el supuesto de la correlación existente entre los tratamientos y las variables explicativas el vector de controles no se puede asumir como determinístico.

4.3. Robustez: empleados con más de 3 años de antigüedad

Es discutible el supuesto utilizado en este trabajo acerca de la exogeneidad en la decisión de sindicalizarse, dada la diferencia en el salario esperado de acuerdo con esta decisión. Sin embargo, si existe ese sesgo, es más probable que suceda con la brecha salarial corriente. Lo que se hace en esta sección es reducir la muestra a los trabajadores que llevan más de tres años en este empleo, de forma que se reduzca la probabilidad de que decisiones de sindicalización dependan de la remuneración de los sindicalizados y no sindicalizados en el presente⁷.

La ecuación 1 contiene tratamientos que dependen de las brechas de remuneración en el pasado; sin embargo, lo que se observa en ese tiempo es el promedio de los ingresos para personas del mismo grupo (sexo, educación, edad) de los sindicalizados y no sindicalizados (también para público y privado). Esta variable es agregada y no hace parte de la determinación del ingreso de la persona en el presente, por lo que así esté correlacionada con el estatus sindical, no afectará la estimación de α_2 (o α_1 para el caso de público y privado) en la misma medida que lo haría si se incluyera toda la muestra.

En las tablas 8 y 9 se muestran las estimaciones de $\delta \Delta x$ y $\Delta 0$ para las personas que llevan más de tres años en ese mismo empleo.

A pesar de utilizar un subgrupo de los trabajadores de la sección 4.1, y de que este subgrupo no fue extraído aleatoriamente, los resultados no cambian significativamente, sobre todo para el caso del ingreso mensual en el 2011, en donde la diferencia total es de 53,2 % para la muestra completa y de 54,9 % para

⁷ No se elimina totalmente la posibilidad de que el trabajador se haya sindicalizado ese mismo año; sin embargo es mucho más probable que si tomó la decisión de sindicalizarse, la hubiese tomado los dos primeros años en este empleo.

Tabla 7. Sindicalizado vs. no sindicalizado con outcome Log. ingreso por hora.

Variable	Diferencia	Dif. explicada	Dif. no explicada	d.e. no explicada	Dif. trat.	Dif. control	P. trat	P. control
Todos	0.1504	0.1150	0.0227	0.0001	-0.0000	0.0128	0.9992	0.6941
Tipo								
Público	0.0424	0.0334	0.0066	0.0002	-0.0000	0.0024	0.9998	0.9654
Privado	0.0857	0.0678	0.0079	0.0001	0.0000	0.0100	0.9980	0.6770
Sexo								
Mujer	0.1705	0.1185	0.0306	0.0001	-0.0000	0.0215	0.9993	0.6453
Hombre	0.1326	0.1087	0.0174	0.0001	-0.0000	0.0066	0.9991	0.7307
Educación								
Ninguna	0.0179	0.0098	0.0071	0.0004	0.0001	0.0009	0.9982	0.5353
Primaria	0.0582	0.0391	0.0151	0.0002	-0.0000	0.0041	0.9985	0.6422
Secundaria	0.0834	0.0578	0.0217	0.0001	-0.0001	0.0041	0.9990	0.7266
Superior	0.0831	0.0377	0.0359	.	-0.0000	0.0095	0.9994	0.8169
Rango de edad								
entre 19 y 24	0.0641	0.0418	0.0174	0.0004	0.0007	0.0043	0.9907	0.5489
entre 25 y 35	0.0957	0.0659	0.0188	0.0002	0.0000	0.0111	0.9996	0.7188
entre 36 y 45	0.1341	0.0961	0.0248	0.0002	0.0000	0.0131	0.9996	0.7185
entre 46 y 55	0.1670	0.1235	0.0298	0.0002	-0.0001	0.0138	0.9994	0.7378
mas de 55	0.2028	0.1669	0.0202	0.0003	-0.0002	0.0158	0.9986	0.6656

Fuente: elaboración propia

Tabla 8. Diferencias del logaritmo del ingreso mensual, personas con más de 3 años trabajando.

Diferencia	2008	2009	2010	2011
Dif. Total	-0,5540***	-0,5522***	-0,5490***	-0,5494***
	(0,0001)	(0,0001)	(0,0001)	(0,0002)
$\Delta 0$	-0,0703***	-0,0684***	-0,0394***	-0,0173***
	(0,0001)	(0,0001)	(0,0001)	(0,0002)
ΔX	-0,4837***	-0,4838***	-0,5096***	-0,5321***
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Obs 133295				

Fuente: elaboración propia

Errores estándar con Bootstrapping con 100 repeticiones. Significancia: * 10 %, ** 5 %, *** 1 %.

Tabla 9. Diferencias del logaritmo del ingreso por hora, personas con más de 3 años trabajando.

Diferencia	2008	2009	2010	2011
Dif. Total	-0,4021***	-0,3776***	-0,3372***	-0,3196***
	(0,0001)	(0,0001)	(0,0001)	(0,0002)
$\Delta 0$	-0,1085***	-0,0980***	-0,0804***	-0,0592***
	(0,0001)	(0,0001)	(0,0001)	(0,0002)
ΔX	-0,2936***	-0,2797***	-0,2567***	-0,2604***
	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Obs 133295				

Fuente: elaboración propia

Errores estándar con Bootstrapping con 100 repeticiones. Significancia: * 10 %, ** 5 %, *** 1 %.

los que llevan más de tres años trabajando. Los signos se mantie-

nen por lo que la brecha sindical de los trabajadores públicos es menor

para todos los años, y en los dos tipos de ingreso.

Conclusiones

Este estudio presenta una primera estimación de la diferencia de los beneficios en términos de remuneración de los trabajadores sindicalizados y no sindicalizados para empleados públicos y privados. Se utiliza la GEIH entre el 2008 y 2011 para los empleados particulares, del gobierno y cuenta propia, sindicalizados y no sindicalizados, entre 19 y 65 años que reporten ingresos por esta actividad mayores a cero. En la muestra utilizada los hombres tienen menor propensión a estar sindicalizados a diferencia de lo que se muestra en estudios tanto internacionales como de Colombia; así mismo, quienes tienen alto nivel educativo presentan mayor probabilidad de estar sindicalizados, lo que es confirmado por la literatura, y los adultos mayores de 36 años se relacionan con mayores tasas de sindicalización.

En todas las especificaciones, los resultados muestran menor respuesta de los salarios al estatus sindical cuando se es empleado público. Adicionalmente, se muestra que la brecha sindical incluyendo los dos grupos es de 15 %, con más incidencia en las mujeres; de igual forma, los trabajadores de mayor edad, así como también los más educados. Se eliminaron los empleados con menos de tres años en el actual trabajo, de forma que la decisión de sindicalizarse esté menos correlacionada con la brecha actual de sindicalización; sin embargo, los resultados son similares, ya que la brecha de sindicalización es menor para los empleados públicos en aproximadamente 35 % cuando se mira la diferencia relativa en el ingreso por hora.

Dentro de los resultados se debe tener en cuenta que posiblemente se están comparando empleados de una misma empresa donde uno está sindicalizado y otro no. Si la empresa firmó una convención colectiva con más de dos terceras partes de los empleados entonces los no sindicalizados recibirán los mismos beneficios, por lo que las estimaciones mostrarían un efecto subestimado de los sindicatos sobre el ingreso. Como extensión a este trabajo se puede eliminar el supuesto de exogeneidad del estatus sindical realizando una corrección de acuerdo con los diferentes tipos de trabajador (sindicalizado privado, sindicalizado público, no sindicalizado privado y no sindicalizado público) similar a como se hizo en Gyourko and Tracy (1986).

Bibliografía

Abowd, J. M. and Farber, H. S. (1982). Job queues and the union status of workers. *Industrial and Labor Relations Review*, 35, 354-367.

Christie, V. (1992). Union wage effects and the probability of union membership. *The Economic Record*, 68, 43-56,

Chrysanthou, G. M. (mayo 2008). *Estimating union wage effects in Great Britain during 1991-2003*. Departamento de Economía. Universidad Carlos III de Madrid.

Cuesta, L. (2005). Impacto de los sindicatos en Colombia: ¿mayores salarios y más desigualdad? *Desarrollo y Sociedad*, 56, 117-219.

Cúellar, M. M. (2009). *Los sindicatos y la asignación del ingreso en Colombia. Un siglo de historia laboral*. Universidad de los Andes. Facultad de Derecho.

Freeman, R. (septiembre 1983). Longitudinal analysis of the effects of trade union. NBER Working Paper Series. Working Paper No. 1207.

Guataquí, J. C., Rodríguez, M. y García, A. (mayo 2009). Determinantes estructurales de la sindicalización en Colombia. Documentos de Trabajo. Universidad del Rosario.

Gyourko, J. and Tracy, J. (may 1986). An analysis of public and private sector wage allowing for endogenous choices of both government and union status. NBER Working Papers. No. 120.

Kuhn, P. (1998). Unions and the economy: what we know; what we

should know. *The Canadian Journal of Economics*, 31(5), 1033-1056.

Landerretche, O., Lillo, N. and Puentes, E. (september 2010). The union effect on wages in Chile: A two-stage approach using panel data.

Mankiw, G. (2000). *Macroeconomía* (4.a ed.).

Nopo, G. (mayo 2008). Matching as a tool to decompose wage gaps. *The Review of Economics and Statistics*, 90(2), 290-299.

Robinson, C. (1984). Union wage differential in the public and private sectors: a simultaneous equation specification. *Journal of Labor Economics*, 2(1), 106-127.

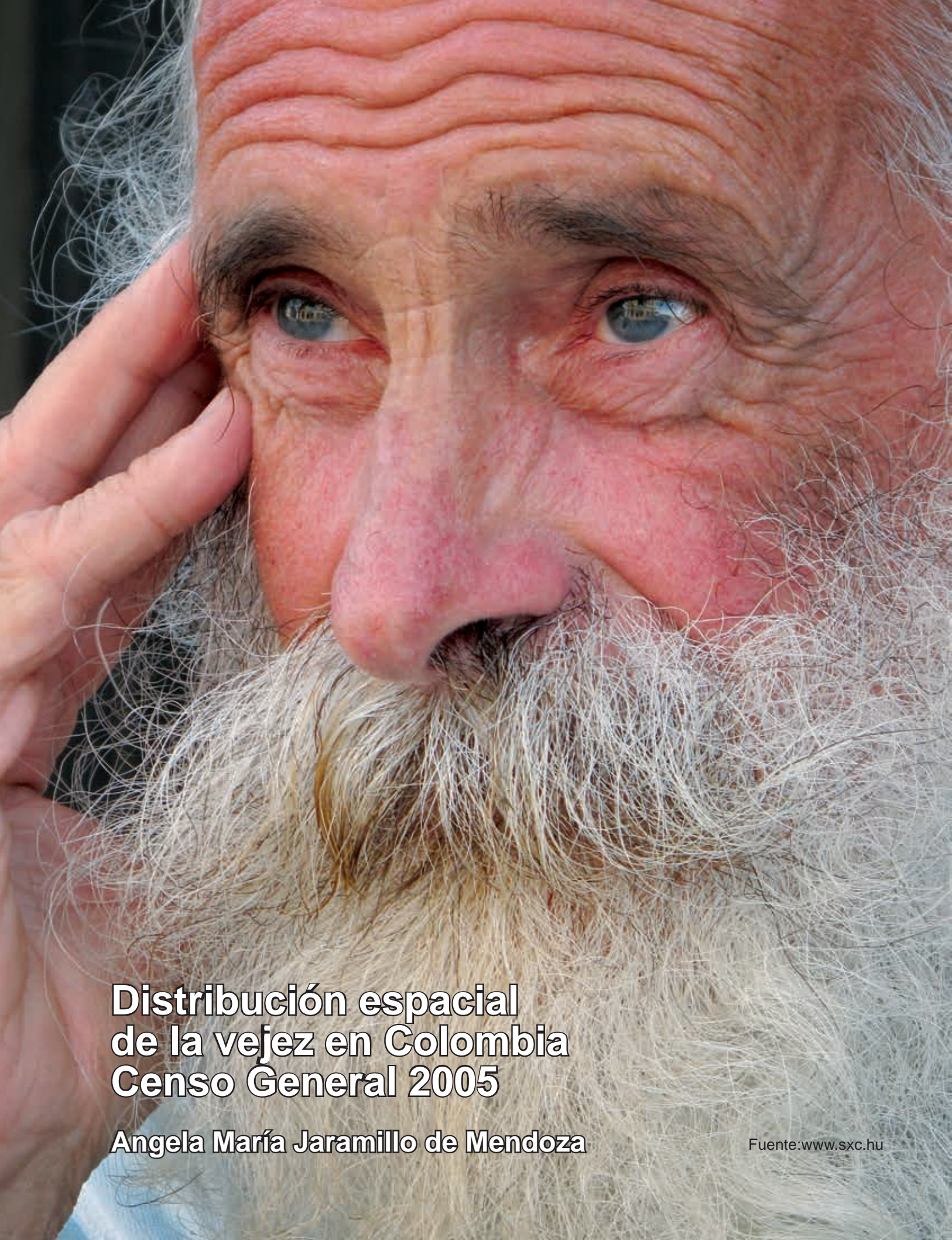
Rodríguez-Lesmes, P., Trujillo, J. D. y Valderrama, D. Más allá de la infraestructura: el impacto de las bibliotecas públicas en la calidad de la educación. Concurso ICFES, 2011.

Schur, L. A. and Kruse, D. L. (1992). Gender differences in attitudes

toward unions. *Industrial and Labor Relations Review*, 46(1), 89-102.

Trujillo, J. D. (noviembre 2011). Intensidad de sindicalización y cambios salariales en Colombia. Tesis de maestría. Universidad de los Andes.

Vázquez, H. (junio 1999). Contenidos y tendencias de la negociación colectiva en Colombia. Área de Negociación, Gestión y Planificación Sindical ENS.



**Distribución espacial
de la vejez en Colombia
Censo General 2005**

Angela María Jaramillo de Mendoza

Fuente: www.sxc.hu

Distribución espacial de la vejez en Colombia Censo General 2005

Angela María Jaramillo de Mendoza: Socióloga de la Universidad Nacional de Colombia, magíster en Estudios de Población de la Universidad Externado y estudiante de Doctorado en Estudios Sociales de la Universidad Externado. Docente investigadora de la Pontificia Universidad Javeriana. Correo electrónico: jaramillo-angela@javeriana.edu.co

Fecha de recepción: 24 de junio de 2013
Fecha de aceptación: 2 de octubre de 2013

Resumen: uno de los principales desafíos socio-demográficos y territoriales que enfrenta el mundo contemporáneo es el envejecimiento demográfico urbano, en el que los asuntos asociados a los entornos ambientales y culturales tendrán un papel cada vez más importante en las políticas sociales y en la planeación urbana. En este documento se considera la relación entre el proceso general de poblamiento en Colombia en el siglo XX y la distribución actual de la vejez.

En este texto, se describen las principales características de las personas de 60 años y más en Colombia y sus departamentos, de acuerdo con su distribución urbano-rural, lugar de nacimiento, hogares particulares, lugares especiales de alojamiento y tipo de vivienda.

Palabras clave: vejez, distribución espacial, generaciones.

Abstract: *one of the main socio-demographic and territorial challenges facing the contemporary world is urban aging population, in which the issues associated with environmental and cultural environments will have an increasingly important role in social policy and urban planning. This paper considers the relationship between the general process of settlement in Colombia in the twentieth century and the current distribution of oldage.*

In this text, we describe the main characteristics of persons 60 years and over in Colombia and its Departments, according to urban-rural distribution, birthplace, private households, special places and type of housing accommodation.

Keywords: *old age, spatial distribution, generations.*

1. Introducción

El envejecimiento de las poblaciones está asociado al lugar y a la vida de la gente, a su historia. Desde finales del siglo XIX hasta comienzos del XX, Colombia registró grandes cambios espaciales, en los que se destaca el despoblamiento y concentración de la Cordillera Oriental, y la formación de la «cuadricéfala urbana» compuesta por Bogotá, Cali, Medellín y Barranquilla (Zambrano y Bernard, 1993). Los desplazamientos de la población en el siglo XX se orientaron especialmente hacia las áreas urbanas, en las que vivía solo el 31% de la población total en 1938, mientras que en 1993 alcanzó el 69%. Este incremento tuvo su mayor expresión hacia la década del cincuenta, en la que se pasó de una tasa de urbanización de 19,5 por mil a 26,1 por mil en 1964, debido especialmente a la explosión demográfica y la migración neta rural urbana (Flórez, 2000).

El avance del proceso de urbanización en el país a lo largo del siglo no fue homogéneo. En la primera mitad de siglo se observaron importantes traslados poblacionales de la Cordillera Oriental a la Cordillera Central, asociados al surgimiento del cultivo del café y sus exportaciones, y nuevas formas de comunicación como el puerto de Barranquilla y el Ferrocarril del Pacífico. Para este momento, las principales ciudades de la zona central del país ofrecían mejores condiciones para la población, un mayor acceso a servicios de educación y salud, así como un mejor suministro de servicios públicos y de transporte.

Entre 1905 y 1918, el país tenía una tasa de crecimiento poblacional baja, con un aumento de las ciudades relativamente estable. Bogotá se destacaba como la principal concentración urbana con 2,5% de la población total. Es a partir de 1938 con las migraciones del cam-

po a la ciudad, la explosión demográfica y la industrialización de Colombia, que se acelera la velocidad de crecimiento de las principales ciudades y se incrementan las concentraciones urbanas. Para 1973, el 40% de la población urbana de Colombia se encontraba en Bogotá, Medellín, Cali y Barranquilla. Luego de 1973, se registra la primacía urbana de Bogotá que concentraba una buena parte de las actividades industriales, bancarias y tecnológicas, con un posterior proceso de metropolización (Flórez, 2000).

La configuración espacial de la vejez en Colombia a comienzos del siglo XXI es el resultado de los procesos de poblamiento de comienzos del siglo XX. Las actuales generaciones envejecidas nacieron entre 1890 y 1945, período en el que se observaron importantes corrientes migratorias internas que se orientaban hacia la concentración de una red urbana, con hegemonía de la capital. En su distribución actual es posible rastrear algunos rasgos de comienzos de siglo como los lugares de origen de esta población, que en su mayoría son distintos a los de su residencia actual, la disminución general de la vida rural, aunque ciertas poblaciones muestran continuidad y permanencia asociada a la baja transformación productiva de algunas zonas del país. Así mismo es posible registrar los cambios habitacionales, ya que para las cohortes que nacieron en este período la forma habitacional más común era la casa, mientras que para las generaciones nacidas luego de los años 60 van a ser los edificios de apartamentos, que responden a nuevas formas de organización socio-espacial, como la densificación de las ciudades y los cambios en la estructura familiar.

Así es como no solo envejece el sujeto en el desarrollo de su curso de vida sino sus entornos humanos y físicos en los que se observan nuevas demandas asociadas

a cada momento histórico. La vejez es el despliegue de unas relaciones sociales construidas a lo largo de la vida que implican unas dinámicas y comportamientos diferentes a los de los colectivos que se encuentran en otros momentos de vida como pueden ser los niños, jóvenes y adultos. Estos modos de vida resultan de la interacción y participación en/con y desde el medio, en el que las circunstancias, condicionantes, permisos y restricciones de las condiciones materiales y simbólicas pasan a primer plano. En las principales ciudades del país se observa que las generaciones hoy envejecidas no muestran mucho interés por migrar a otras localidades, comunas o barrios posiblemente por el aprecio que tienen por sus barrios y casas como parte de su historia y sentido de vida. Mientras que las generaciones más jóvenes pueden estar mostrando mayor movilidad orientada por sus expectativas de vida y su mayor familiaridad con el cambio como forma de vida. Sus relaciones con los lugares son múltiples y heterogéneas, ya que dependen de las rutas de vida de los sujetos que componen cada cohorte. Estos actores van configurando una multiplicidad de territorialidades, entendidas como formas de apropiación y reproducción del entorno, que es posible reconstruir para comprender la naturaleza de sus comportamientos presentes, con el fin de diseñar políticas públicas que faciliten la construcción de ciudades para todas las edades.

2. Resultados

Según el Censo de 2005, en Colombia el 75,0% de las personas de 60 y más años vive en cabeceras, un punto por debajo del total nacional. El reto del envejecimiento implica una planeación urbana gerontológica que resulte de la colaboración de equipos multidisciplinares (ciencias

de la salud, ciencias sociales, arquitectura, ingenierías, entre otros), con la intención de crear condiciones de habitabilidad (viviendas, barrios, bienes y servicios urbanos en general) necesarias para una población que envejece.

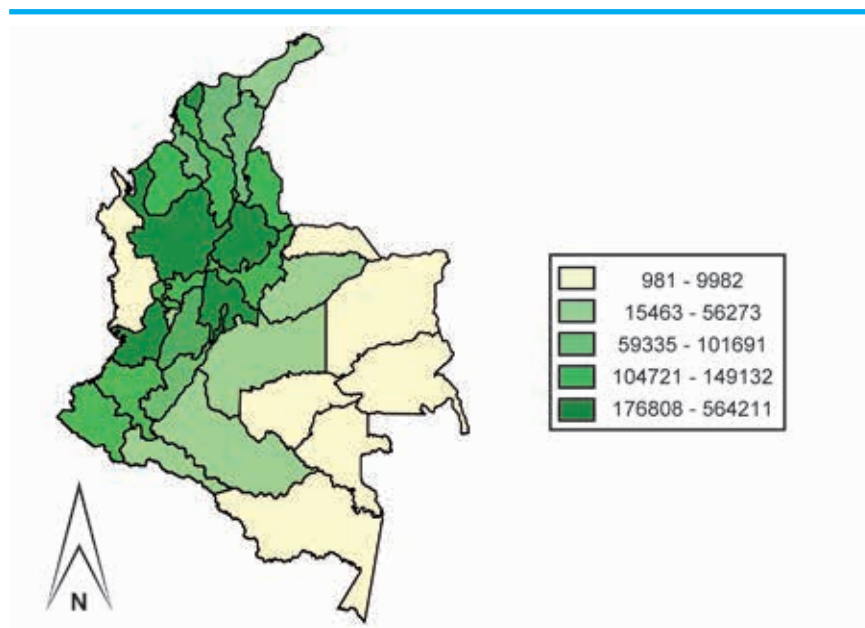
Los grupos humanos no envejecen en abstracto, lo hacen en tiempos y espacios concretos, en los que han desarrollado sus historias de vida. La población es la que construye y le da sentido al espacio que habita, «*Los objetos del espacio geográfico, son resultados de la relación directa o indirecta del trabajo humano, simples o complejos mediados por la intencionalidad, la significación, la técnica y la estética*» (Montañez, 1999). La organización espacial, en tanto territorio, va más allá de su condición de Estado, es histórica, funcional y relacional, y sugiere un conjunto de relaciones de dominio, pertenencia, identidad y apropiación entre un espacio geográfico y un sujeto individual o colectivo (Agnew, 2008: 15-30). Para comprender el comportamiento y las dinámicas de las poblaciones envejecidas, es necesario estudiar su relación con los entornos ambientales y culturales que condicionan sus características, experiencias y calidad de vida. En este documento se identifican algunos aspectos espaciales de la población envejecida en Colombia en 2005, como su distribución, departamentos de nacimiento, tipo de alojamiento y tipo de vivienda.

3. Distribución urbano-rural

En el año 2005¹, en Colombia vivían 3.721.943 personas de 60 y

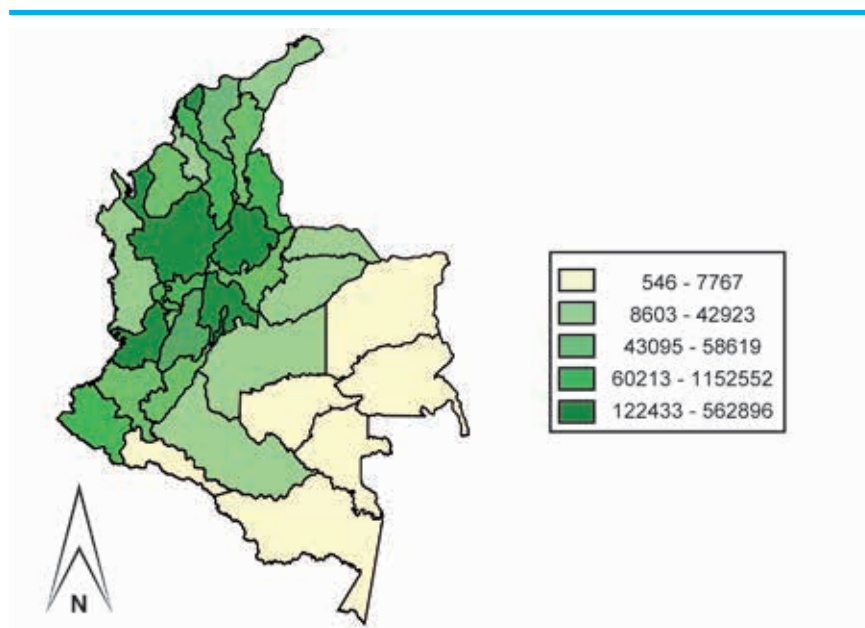
¹ Los análisis que se presentan en este texto son en valores absolutos y en relación con el total de la población mayor, por lo que no se puede concluir que estos sean los territorios más envejecidos del país, ya que para eso hay que calcular los porcentajes de esta población respecto a los totales de sus municipios, en cuyo caso los departamentos que tienen más participación de las personas mayores son Boyacá, Caldas, Tolima y Quindío. Sin embargo, los mapas disponibles en el sistema de consulta del DANE no permiten referenciar esta información.

Mapa 1. Ambos sexos. Número de PM según departamento - 2005.



Fuente: DANE Censo General 2005.

Mapa 2. Ambos sexos. Número de PM según cabecera - 2005.

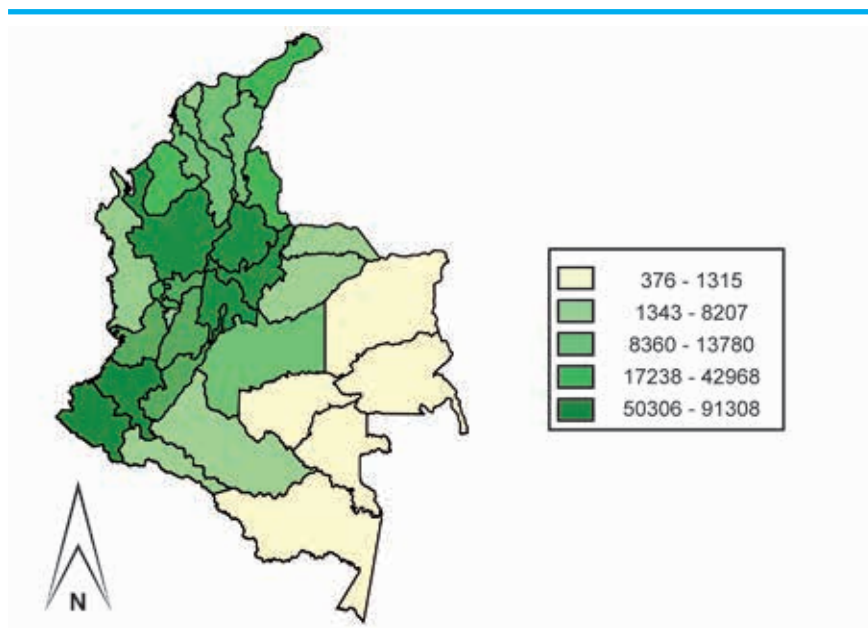


Fuente: DANE Censo General 2005.

más años (46% hombres y 54% mujeres). La mayoría (55,6%) de estas personas se encontraron en la región central del país (Bogotá con 564.211 y Cundinamarca con 222.434), seguida por Antioquia con 520.599, Valle del Cauca con 399.090, Santander con 186.105 y Atlántico con 176.808 (mapa 1).

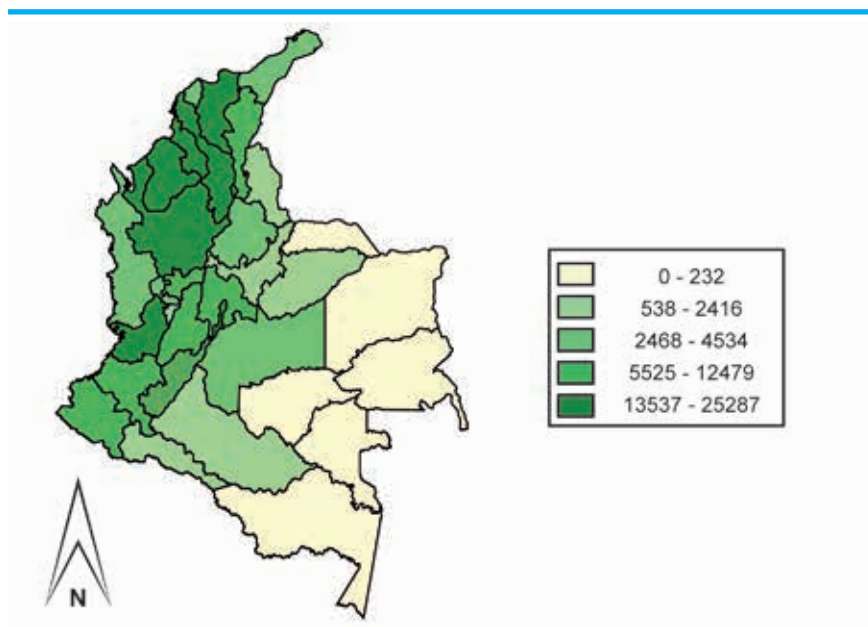
Como se mencionó, el 75% del total nacional se encontró en las cabeceras, especialmente en Bogotá, Antioquia y Valle del Cauca (mapa 2), mientras que el 19,6% en zonas rurales como las de Antioquia, Boyacá, Cauca, Cundinamarca, Nariño, Santander y Tolima (mapa 3), y solo un 5,4% en centros poblados,

Mapa 3. Ambos sexos. Número de PM según rural disperso - 2005.



Fuente: DANE Censo General 2005.

Mapa 4. Ambos sexos. Número de PM según centro poblado - 2005.



Fuente: DANE Censo General 2005.

especialmente en la región costera con municipios como Córdoba, Magdalena y Bolívar (mapa 4).

La distribución espacial de esta población refleja el proceso de poblamiento del país desde mediados del siglo XIX, con el surgimiento de numerosas ciudades en la Cordillera

Central, que comenzaron a disputar la primacía que desde épocas prehispánicas tenía la Cordillera Oriental en el eje Bogotá-Pamplona. Este cambio se produjo como resultado de las transformaciones en la economía. Las tierras altas de la Cordillera Oriental se convirtieron a finales del XIX en zonas expulsoras

de población que era recibida por los nuevas tierras templadas donde se cultivaba café, y por las ciudades puerto, proceso que se va a consolidar hacia los años cuarenta del siglo XX, con el surgimiento de los grandes centros urbanos: Bogotá, Bucaramanga y Cúcuta, seguidos de Cali, Barranquilla y Medellín, rodeados de ciudades intermedias que se conservan en una relación funcional con ellas. Finalizando el siglo XX el panorama es similar al que se registró 200 años antes: el Altiplano Cundiboyacense y la Costa Atlántica son las regiones más pobladas del país, acompañadas por el Valle del Cauca y las nuevas zonas de colonización. Una excepción de esta nueva primacía urbana es Bogotá, que conservó su importancia en el largo plazo (Zambrano y Bernard, 1993). Las generaciones hoy envejecidas en el país nacieron entre 1890 y 1945, cuando se estaba consolidando la cuadricefalia urbana: Bogotá, Medellín, Cali y Barranquilla, que organizó al país en cuatro grandes regiones que reflejan las realidades histórico-culturales del largo plazo.

4. Departamento de nacimiento

Cuando se llega a los 60 años de vida, la persona cuenta con una amplia y variada acumulación de experiencias; una de ellas es la residencial, en la que se puede encontrar una alta o baja movilidad según los cursos biográficos que cambian de acuerdo con las condiciones históricas de los sujetos. Para las generaciones nacidas en la primera mitad del siglo XX, moverse era una tarea muy costosa en tiempo y dinero; hasta los años veinte las poblaciones se transportaban por lo general en mulas o carros de tracción animal por carreteras en mal estado o caminos de herradura, mientras se avanzaba en la ampliación del ferrocarril que había llegado al país desde 1870 como complemento

de las vías fluviales (Magdalena y Cauca). Esto ocasionaba altos costos para los traslados y promovía el aislamiento de las regiones y los mercados fragmentados. Tales limitaciones condicionaron la experiencia de movilidad de estas generaciones, adicionalmente los desarrollos económicos de las regiones influían en la expulsión o atracción de estas poblaciones.

4.1. Región central (Bogotá y Cundinamarca)

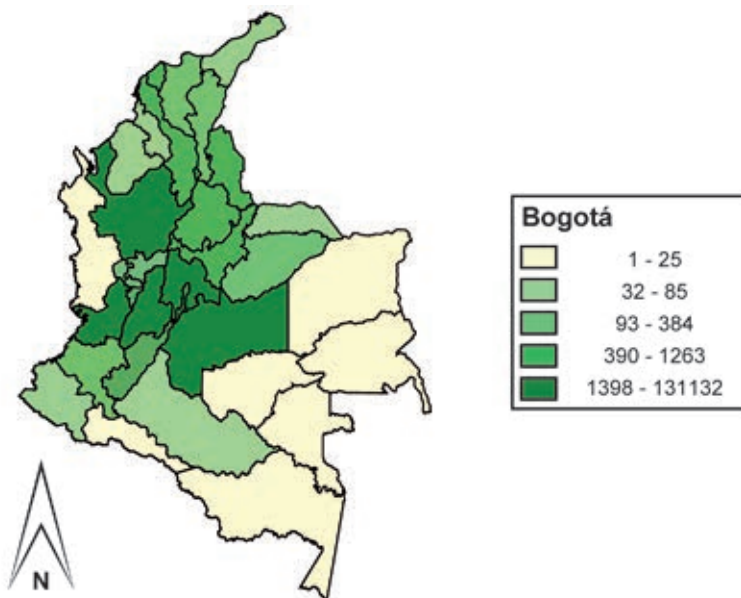
Cerca de la mitad de las personas que viven en Bogotá nacieron en Boyacá (20,8%) y Cundinamarca (25,1%), solo el 24,2% nació en Bogotá² (mapa 5), mientras que el 72,8% de los habitantes de Cundinamarca nacieron allí mismo. Esto puede estar asociado a que el 41% (91.308) de los pobladores de Cundinamarca viven en la zona rural dispersa, de los cuales el 81% (76.064) nació allí. Mientras que en Bogotá solo 1.305 (0,2% del total) viven en zona rural (mapas 5 y 6).

Entre 1800 y 1907 la población de Bogotá se multiplicó por cinco; su crecimiento fue producto especialmente de los procesos de inmigración, ya que para esa época la mortalidad por epidemias y guerras era muy alta (Mejía, 1999). La ganancia poblacional de la ciudad se debía en buena parte a su tradicional importancia como centro político-administrativo; allí vivía un buen segmento de las élites del país que desde 1870 venían acumulando riquezas que eran producto de las bonanzas exportadoras. Para ese momento la ciudad se estaba densificando como resultado de su crecimiento poblacional y de la diversificación económica de los usos del suelo, entre los que se destacan los bancos, las casas de comercio, las agencias de negocios, las universidades, centros de salud, restaurantes, entre otros; además de sus cambios en infraestructura vial con el tranvía y el automóvil, así como el surgimiento de los servicios públicos de luz y agua, acompañados de innovaciones en comunicación como la radio y el cinematógrafo. Bogotá no vivía

directamente los procesos de industrialización que se observan en otros departamentos como Antioquia y Santander, pero su centralización financiera y política, así como sus innovaciones tecnológicas constituían un atractivo para las poblaciones vecinas que veían en la ciudad nuevas oportunidades para trabajar y mejorar sus condiciones de vida, especialmente los pobladores de los departamentos de Boyacá y Cundinamarca en los que no se registraban mayores transformaciones en sus formas de producción campesina que se caracterizaban por un etos minifundista y de subsistencia.

Es posible que la alta permanencia que se observa en la población envejecida de Cundinamarca se relacione con la baja industrialización y modernización del territorio que no provocó significativos cambios residenciales en su población, conservando una distribución de predominio rural en el que es posible que se conserven buena parte de los significados y sentidos de la producción de subsistencia de 100 años atrás. Este es el único departamento que registró una baja urbanización de la población envejecida, y se destaca por la proporción de personas de 60 y más que no recibieron educación y no saben leer y escribir. Lo que puede mostrar cómo la oferta educativa que se concentró en las ciudades estableció mayores distancias con las zonas rurales que no tenían un sistema educativo orientado a mejorar las formas de producción campesina y la calidad de vida de esta población. Esto es algo que todavía se observa en los sitios más alejados de los departamentos, lo que reproduce las brechas entre las poblaciones urbanas y rurales y sus condiciones de vida.

Mapa 5. Bogotá. Número de PM según departamento de nacimiento - 2005



Fuente: DANE Censo General 2005.

² Una proporción muy baja si se compara con la población total de Bogotá, en la que el 87% nació en esta misma ciudad.

³ Es baja esta proporción si se compara con el total de la población del departamento, de la cual el 77% nació allí mismo.

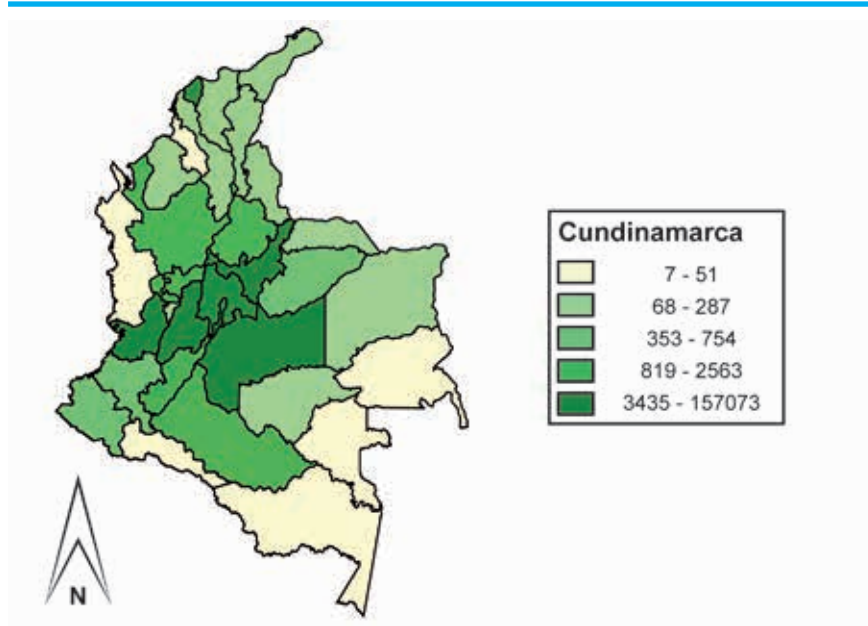
4.2. Valle del Cauca y Atlántico

Por su parte el Valle del Cauca muestra una mayor diversificación frente a los dos departamentos mencionados. La mitad³ de sus poblado-

res (52,5%) nacieron en el mismo departamento, y la otra mitad proviene especialmente de Cauca (8,1%),

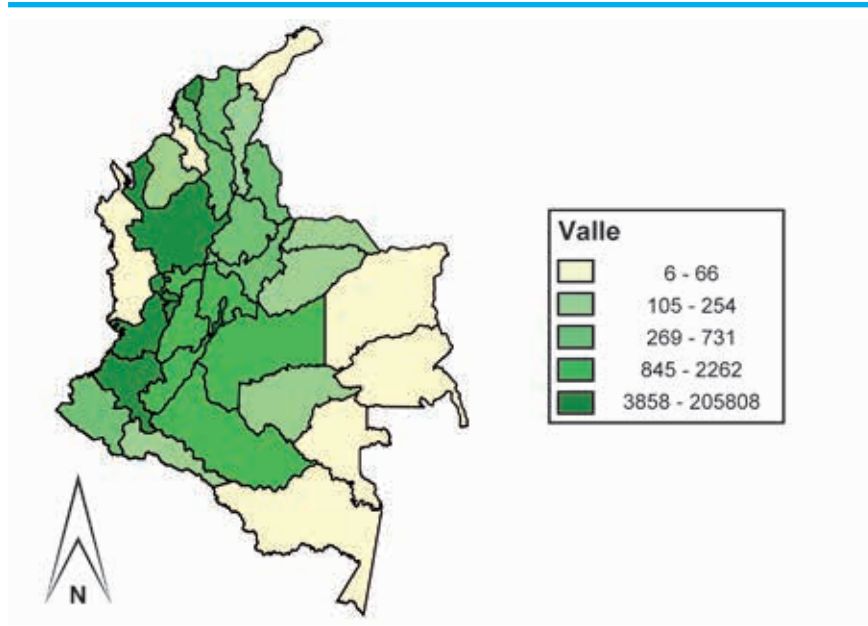
Nariño, Caldas (6,9%), Antioquia (4,8%), Tolima (4,6%), Risaralda (3,9%) y Quindío (3,6%) (mapa 7).

Mapa 6. Cundinamarca. Número de PM según departamento de nacimiento - 2005.



Fuente: DANE Censo General 2005.

Mapa 7. Valle del Cauca. Número de PM según departamento de nacimiento - 2005.



Fuente: DANE Censo General 2005.

Similar al Atlántico, con 61,5% de habitantes que nacieron allí. Y el resto de sus habitantes vienen de los departamentos de Bolívar

(11,9%), Magdalena (10,6%) y Sucre (3,8%), principalmente (mapa 8). Esta distribución de la población puede estar asociada a que ambos

departamentos surgieron a comienzos del siglo XX; las poblaciones que hoy son del Valle y del Cauca pertenecían al antiguo Estado del Cauca. Lo mismo sucedió con las poblaciones de Atlántico y Bolívar. Lo que muestra que la distribución actual no se puede explicar solamente por procesos de migración sino por las reconfiguraciones político-administrativas de comienzos de siglo. Especialmente porque estos departamentos no tuvieron en ese momento procesos económicos tan significativos como los que se observaron en Antioquia y Santander. Sin embargo es interesante la composición del Valle porque es más diversificada ya que tiene población de Nariño, Antioquia y el Eje Cafetero, los dos últimos como resultado de la colonización; mientras que Atlántico no registró población de otras regiones del país.

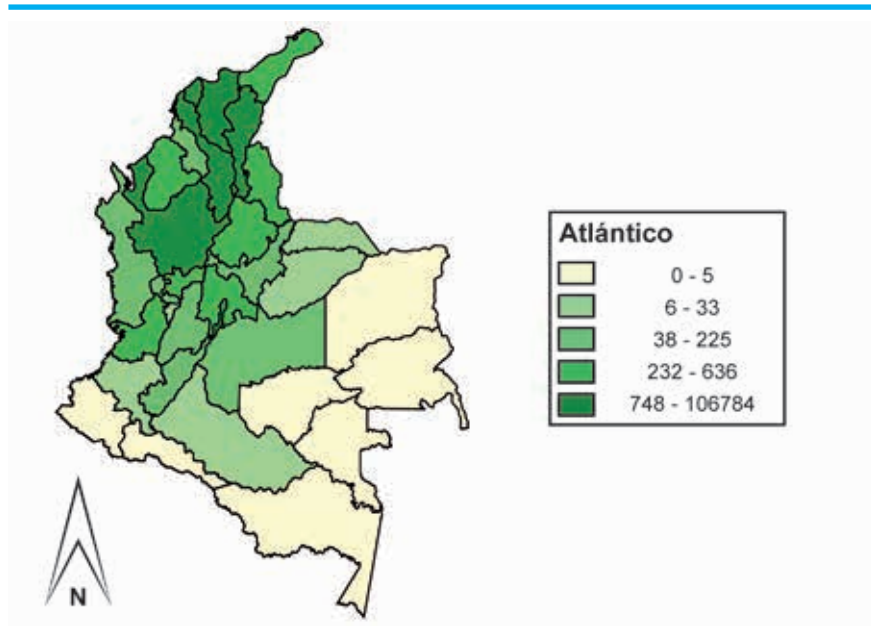
4.3. Antioquia

A diferencia de Bogotá, Valle del Cauca y Atlántico, en Antioquia la gran mayoría (86,9%) de las personas de 60 y más años nacieron allí, como en Cundinamarca (mapa 9). Sin embargo, a diferencia de Cundinamarca, solo el 16,8% (72213) vive en la zona rural y 79,7% (343741) en las cabeceras. Lo que refleja los cambios económicos que experimentó el departamento como resultado de la distribución de una propiedad territorial más amplia como el latifundio, en contraste con el centro del país con el minifundio, de los avances en el sector minero, la colonización del sur, los ingresos por las exportaciones del café y la inversión en industria, especialmente en Medellín, lo que movilizó a las poblaciones de las zonas rurales a las urbanas (mapa 9).

4.4. Santander

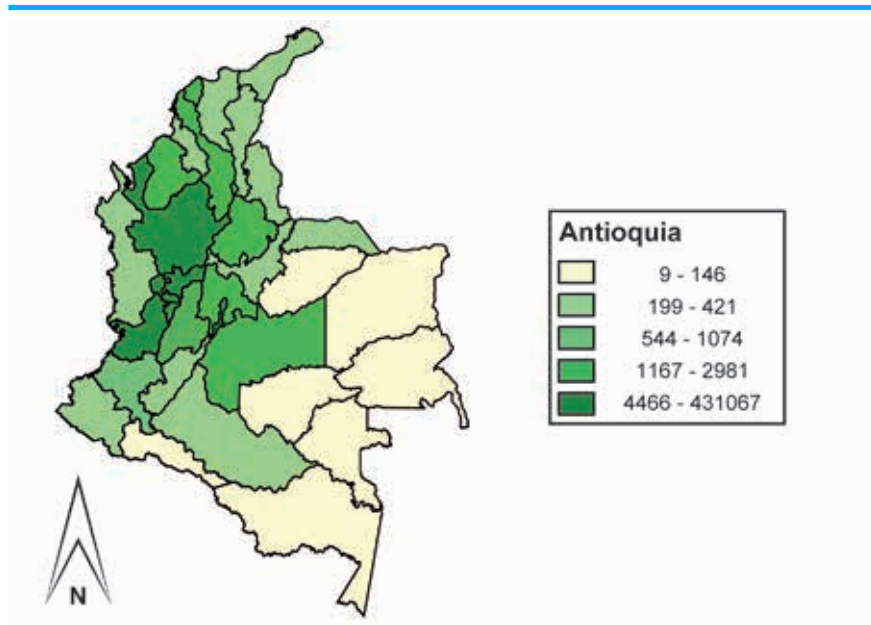
En el caso de Santander también se observa que la gran mayoría de su población (82,1%) nació allí mismo. Con una población urbana de

Mapa 8. Atlántico. Número de PM según departamento de nacimiento - 2005



Fuente: DANE Censo General 2005.

Mapa 9. Antioquia. Número de PM según departamento de nacimiento - 2005



Fuente: DANE Censo General 2005.

72,8% (134.207), también refleja un importante proceso de urbanización e industrialización que se vivió en gran parte por la actividad petrolera del siglo XX con empresas como la Tropical Oil Company y luego Ecopetrol en Barrancabermeja, así

como con la industria manufacturera en Bucaramanga (mapa 10).

En general, se observa que los departamentos de Antioquia, Santander y Cundinamarca registran una alta permanencia de sus poblacio-

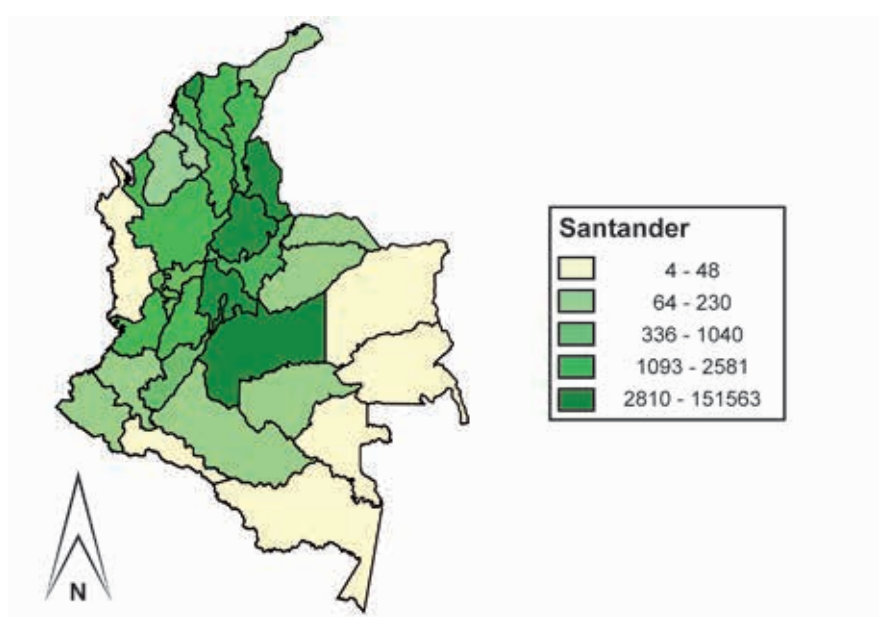
nes a nivel departamental. Esto puede estar asociado a un mayor arraigamiento cultural: «En la medida que aumenta la duración de la residencia en un lugar, el individuo va estableciendo vínculos con el entorno, que se fortalecen con la permanencia en el mismo; las actividades cotidianas van delimitando un espacio de vida, de forma paralela al establecimiento de una serie de redes sociales; todo ello contribuye a arraigar al sujeto a un determinado territorio, dificultando la toma de una decisión de movilidad» (Puga, 2004).

Así mismo, las circunstancias del lugar de nacimiento influyen positivamente en el futuro migratorio de las personas, por lo que es probable que estas generaciones en su mayoría se movieron hacia las áreas urbanas cercanas a sus lugares de origen porque no vieron la necesidad de desplazarse a otros departamentos en el país, especialmente en los casos de Antioquia, Santander y Cundinamarca. Mientras que en los departamentos de Atlántico y Valle del Cauca cerca de la mitad de la población que nació allí migró a otros departamentos; probablemente las condiciones económicas de su lugar de origen no presentaban la misma atracción que las registradas en los territorios mencionados. Adicionalmente, el departamento del Valle se crea en 1910, lo que también puede estar asociado a movimientos poblacionales que reflejaban las contradicciones políticas, territoriales y económicas del antiguo Estado del Cauca, factores que probablemente expulsaron población nativa y atrajeron otras poblaciones. El departamento del Atlántico también se crea en la primera década del siglo XX. En contraste con Bogotá, Cundinamarca, Antioquia y Santander, que son de las entidades territoriales más antiguas del país.

Las personas que nacieron en las cabeceras no tuvieron la misma ne-

cesidad de moverse a lo largo de su curso de vida, probablemente porque allí tenían trabajo, y los servicios «modernos» de salud y educación. Son, por el contrario, los nacidos en áreas rurales los que presentan una trayectoria migratoria más activa, a excepción de Cundinamarca que registra la permanencia de buena parte de su amplia población rural. La estabilidad en el entorno geográfico y social reduce las opciones de optar en la vejez por una movilidad por la que no se decidieron a edades más jóvenes; así lo muestra el 85% de la población mayor que no cambió de residencia en los 5 años antes del censo, mientras que el 15% que se movió lo hizo principalmente por asuntos familiares, por trabajo y por salud.

Mapa 10. Santander. Número de PM según departamento de nacimiento - 2005.

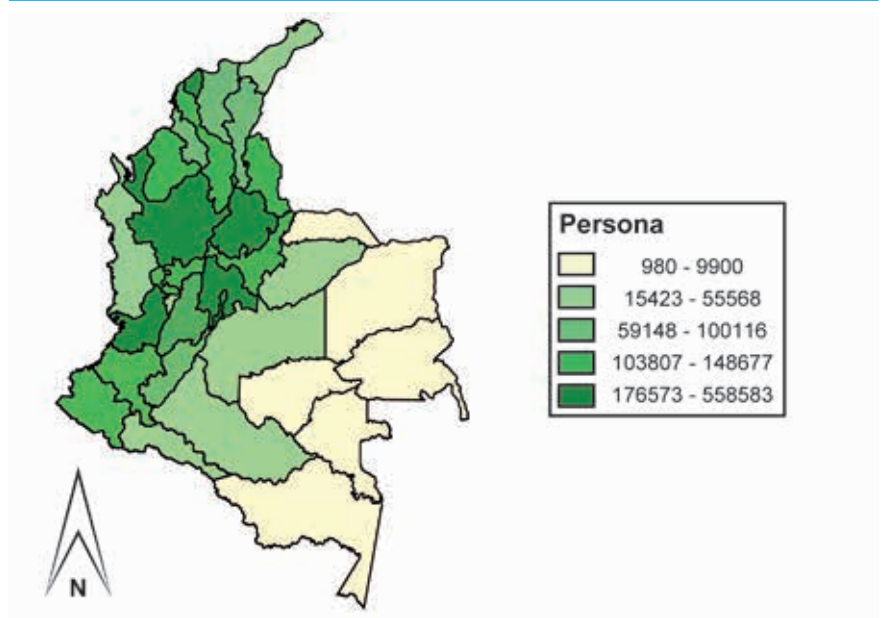


Fuente: DANE Censo General 2005.

5. Hogares particulares y lugares especiales de alojamiento

De los 3.721.943 de personas mayores, el 99,01% vive en hogares particulares (mapa 11); solo el 0,99% en Lugares Especiales de Alojamiento (LEAS). Los departamentos que registran más presencia de los LEAS son: Antioquia, Bogotá, Cundinamarca, Boyacá, Caldas, Santander, Risaralda y Valle del Cauca (mapa 12 y gráfico 1). Es interesante ver que los lugares donde hay más personas institucionalizadas no se corresponden totalmente con los que tienen personas de 100 y más años, ya que se esperaba que la mayoría de los centenarios estuvieran en instituciones por sus pérdidas físicas y funcionales, sin embargo solo 151 (4,7%) de los 3.165 centenarios que tenía Colombia en 2005 estaban institucionalizados. Los departamentos de Córdoba, Bolívar y Magdalena muestran una importante presencia de personas de 100 y más años, que en su gran mayoría viven en hogares particulares (mapas 13 y 14).

Mapa 11. Colombia. Número de PM según hogares particulares - 2005.



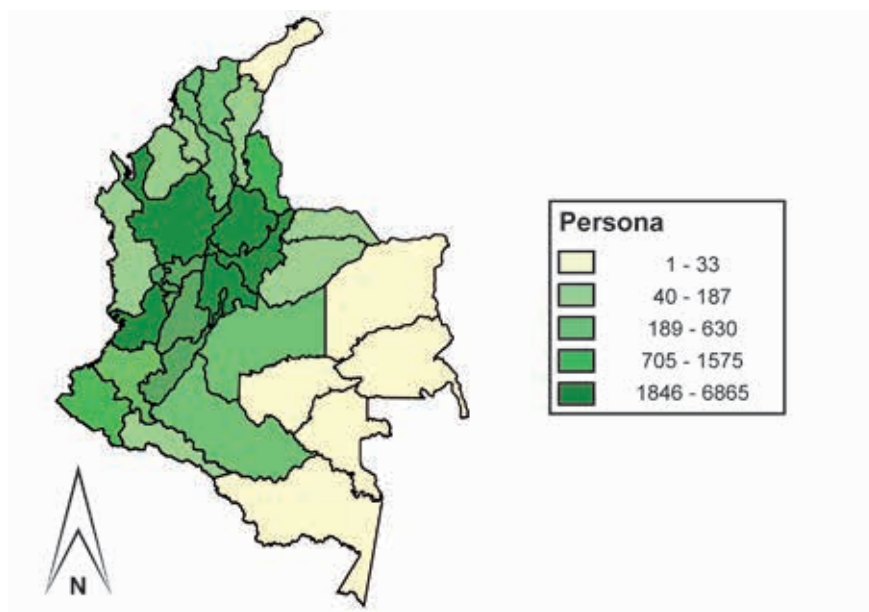
Fuente: DANE Censo General 2005.

6. Vivienda

Los usos y significados de la vivienda van cambiando a lo largo de la vida. En la primera infancia y la vejez es cuando más tiempo se pasa en el espacio doméstico, ya que las actividades en estas

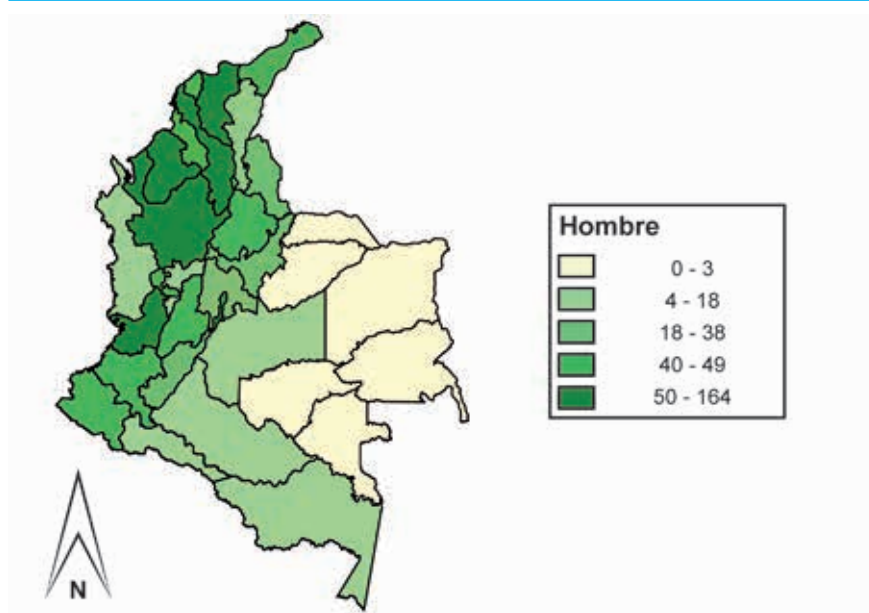
etapas de la vida no están condicionadas por el estudio o el trabajo. Sin embargo, este descenso de la actividad en la vejez es relativo, pues depende de los entornos socio-económicos en los que se realice. En Colombia, para el año 2005, el 20% de la población mayor del país trabajó (80% hombres y 20%

Mapa 12. Colombia. Número de PM según Lugares especiales de alojamiento (LEA) - 2005.



Fuente: DANE Censo General 2005.

Mapa 13. Hombres. Número de PM según edad (100 o más años) - 2005.



Fuente: DANE Censo General 2005.

mujeres), mientras que una tercera parte (33%) se dedicó a las labores del hogar (91% mujeres y 9% hombres), 7% estaba incapacitado para trabajar y 15% vivió de la pensión (60% hombres y 40% mujeres).

El 76% de las personas mayores viven en casas, 7 puntos por enci-

ma del promedio de la población total; y solo el 20% en apartamentos, 5 puntos por debajo del total nacional (mapa15). Esta distribución se explica porque durante la primera mitad del siglo XX, las casas de patio y de fachada fueron las que tejieron nuestras ciudades. Desde la época de la colonia hasta comien-

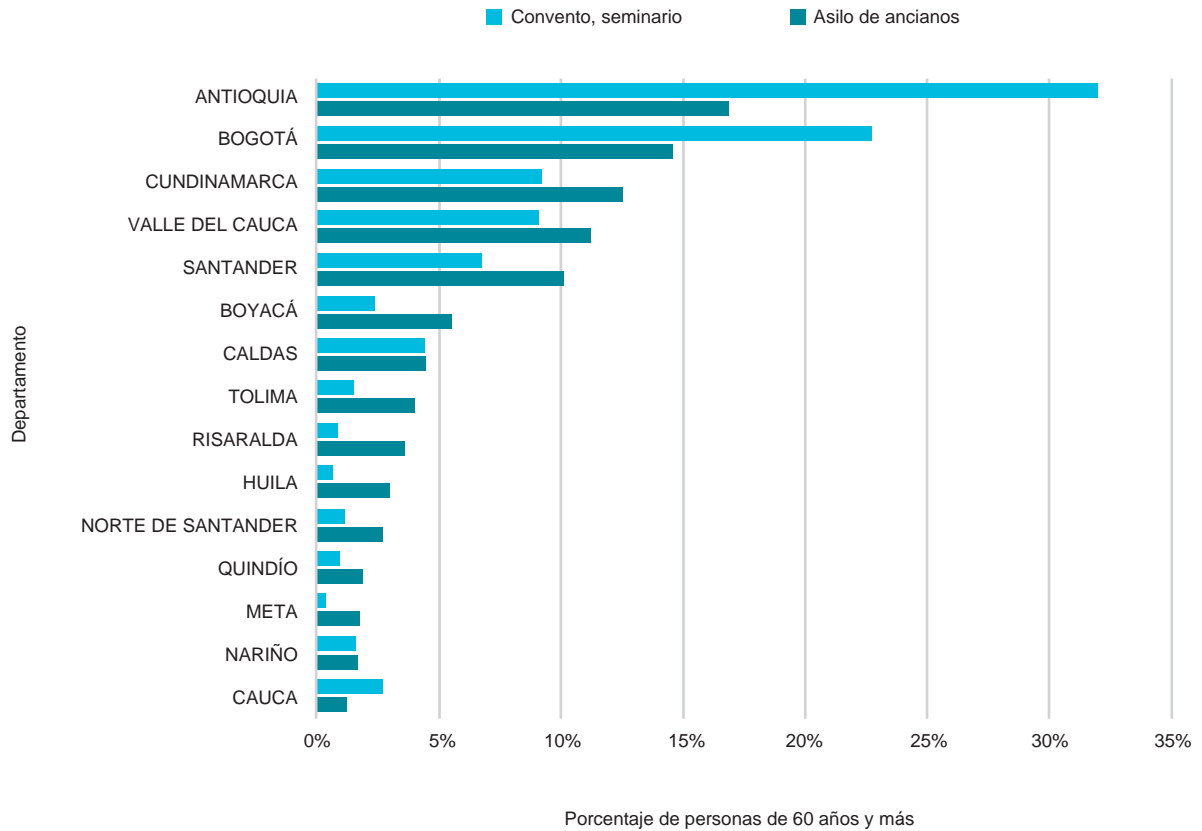
zos del siglo XIX el ordenamiento espacial de las poblaciones estaba orientado por las parroquias, a las que le siguieron los barrios en la primera mitad del siglo. Es hasta la década del cincuenta que surgen los edificios de apartamentos que acompañan los procesos de aglomeración y densificación urbanas (Arango G., Peláez P. y Wolf G., 1995). Estas formas de habitabilidad también son evidencia de las maneras de organización familiar en las que se destacan los grupos extensos y nucleares, orientados por la pareja y sus hijos.

En casa indígena solo se encontró el 1% (32.765), especialmente en los departamentos de La Guajira, Cesar, Córdoba, Chocó, Valle del Cauca y Nariño (mapa 16). Mientras que en cuartos viven 100.259 personas (3%, menor al total nacional: 5%), principalmente en Atlántico, Antioquia, Cundinamarca, Tolima y Valle del Cauca (mapa 17).

Como se mencionó inicialmente, una de las prioridades para que la sociedad pueda adaptarse adecuadamente a su envejecimiento es pensar en las condiciones de habitabilidad de todas las edades. En la medida que avanza la edad, el cuerpo cambia y reduce relativamente sus habilidades físicas, por lo se deben adecuar los espacios domésticos para garantizar accesos confortables, puntos estratégicos de apoyo, ubicación de espacios y objetos de importancia a cierta altura, luminosidad, y los espacios públicos para desarrollar actividades de esparcimiento y asociación.

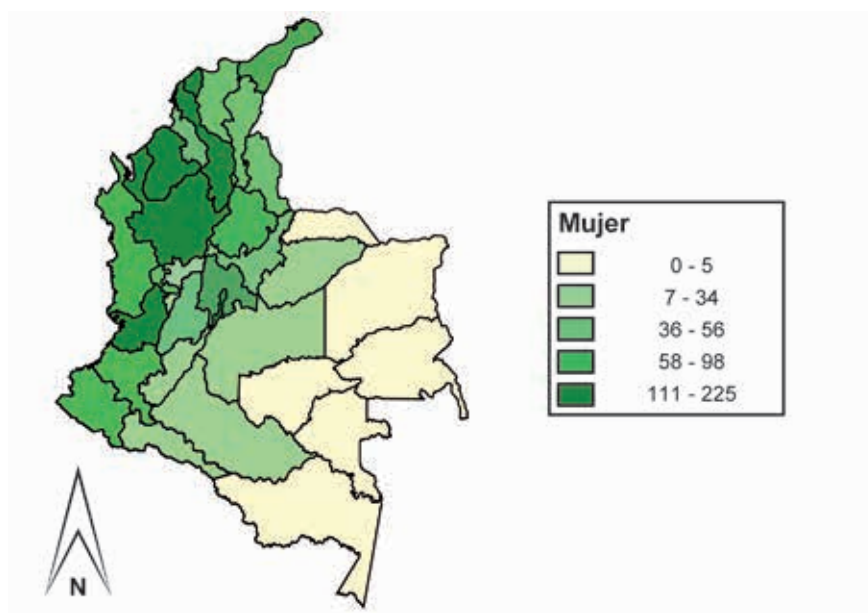
También se puede pensar en mejores accesos a servicios públicos, para sus pagos y trámites, que no impliquen largos trayectos con servicios de transporte inadecuados que no respondan a las necesidades de las personas mayores; esto requiere una planificación local que se cualifique a través del conocimiento de sus habitantes mayores,

Gráfico 1. Departamentos. Porcentaje de personas de 60 años y más según LEA (convento y asilo de ancianos) - 2005.



Fuente: Censo General 2005.

Mapa 14. Mujeres. Número de PM según edad (100 o más años) - 2005.

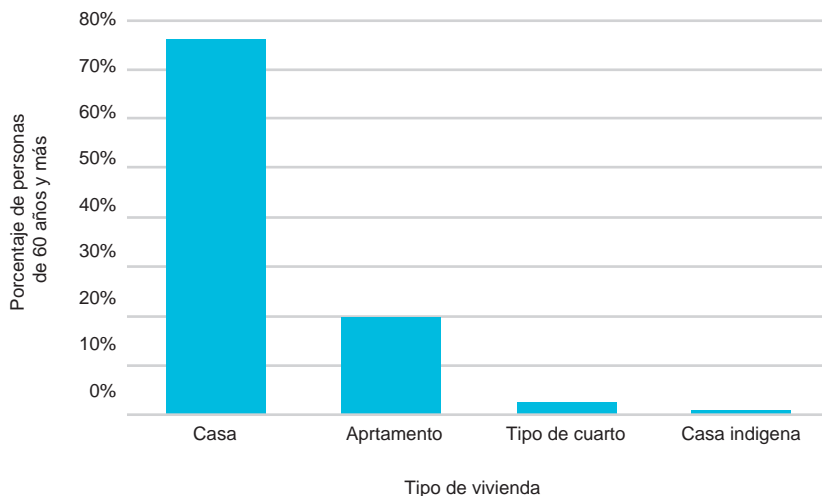


Fuente: DANE Censo General 2005.

y de cómo, según las generaciones, van cambiando sus habilidades para vivir en sus grupos sociales. Así mismo es relevante considerar que las condiciones de habitabilidad no se limitan a lo físico, pues las personas tienen unos entornos afectivos que reflejan sus redes de apoyo; el mantenimiento y la mejora de estas redes también es un asunto de planeación, ya que con la disminución de la fecundidad y los procesos de subjetivación, las redes tradicionales tienden a desaparecer y, en consecuencia, los apoyos que estas garantizaban.

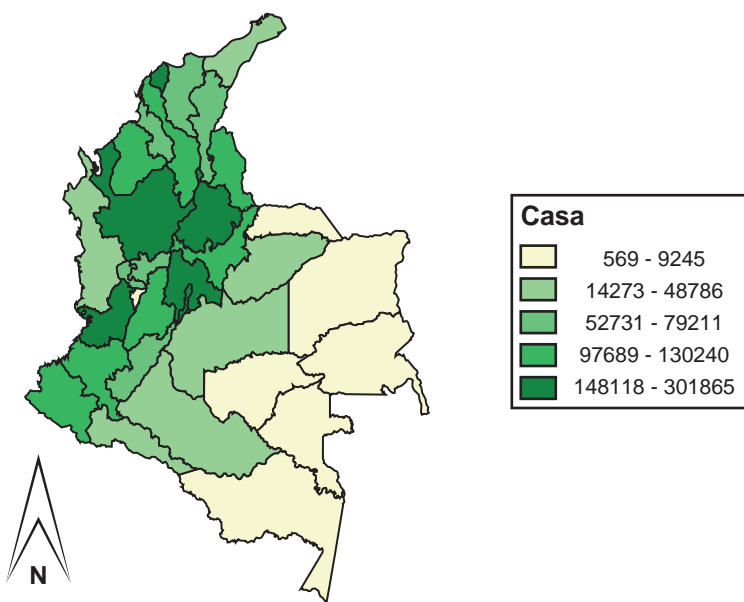
No todas las personas envejecen igual; en Colombia el 20% de la población mayor registró limitaciones, mientras que el 80% no. Sin embargo, todas desarrollan sus actividades

Gráfico 2. Colombia. Porcentaje de personas de 60 años y más según tipo de vivienda - 2005



Fuente: DANE Censo General 2005.

Mapa 15. Departamentos. Número de PM según tipo de vivienda (casa) - 2005.



Fuente: DANE Censo General 2005.

cotidianas en espacios domésticos y públicos, por lo que es necesario crear las condiciones para que tanto las personas que llegan a sus últimos años en condiciones físicas y mentales adecuadas como las que llegan con limitaciones puedan habitar sus lugares en equidad.

7. Síntesis

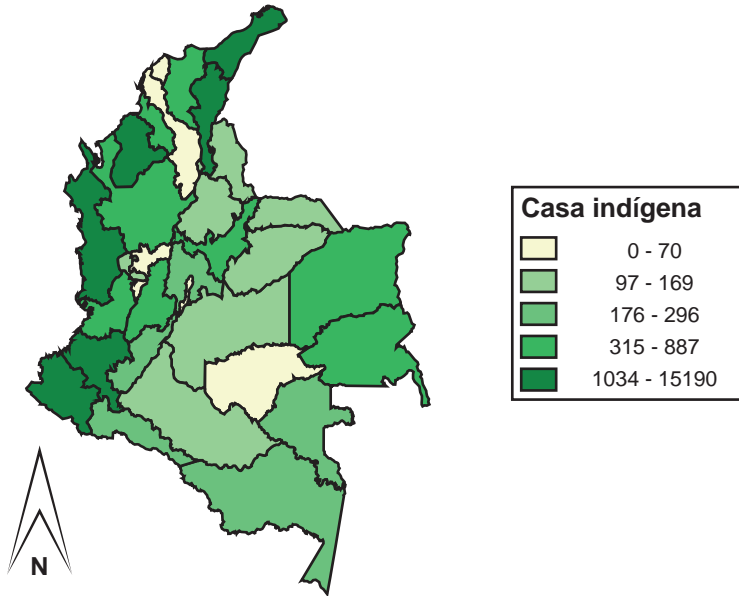
El estudio de la distribución espacial de la vejez es una manera de aproximarse a la relación ecológica que se establece entre los sujetos desde que nacen hasta que mueren y los entornos socio-históricos

en los que se desenvuelve su experiencia social. Esta relación con los lugares se construye de manera heterogénea, según las rutas de vida individual y colectiva, las cuales dependen de las formas históricas de producción material y simbólica de los grupos humanos.

Las actuales generaciones viejas en el país nacieron entre 1895 y 1950, por lo que sus formas de acumulación y evolución vital, así como la conformación de sus entornos sociales, se establecieron en relación con los grandes cambios socio-económicos registrados durante el siglo XX en Colombia. Entre tales cambios se destacan: la emergencia de una economía cafetera, la guerra de los mil días, la construcción de vías de comunicación férreas y de navegación que conectan las principales regiones del país con el extranjero, y la creación de las bases para el auge exportador de comienzos del siglo XX; así como las transformaciones socio-productivas que se reflejan en los procesos de industrialización, cambio de la estructura del empleo, urbanización, endeudamiento externo y la modernización del Estado, entre otros.

En términos generales, las trayectorias espaciales de la población mayor de comienzos del siglo XXI son un registro que permite conocer el proceso general de poblamiento y urbanización del país, en el que se observaron importantes corrientes migratorias internas que estuvieron orientadas por los cambios productivos y de infraestructura que facilitaron la movilidad poblacional en la primera mitad del siglo XX. Las poblaciones mayores de 80 años ubicadas actualmente en Bogotá, Valle del Cauca y Atlántico corresponden a las primeras generaciones de migrantes rurales que conformaron los principales centros urbanos de estos departamentos. Ellos presentaron una mayor diversificación de los lugares de

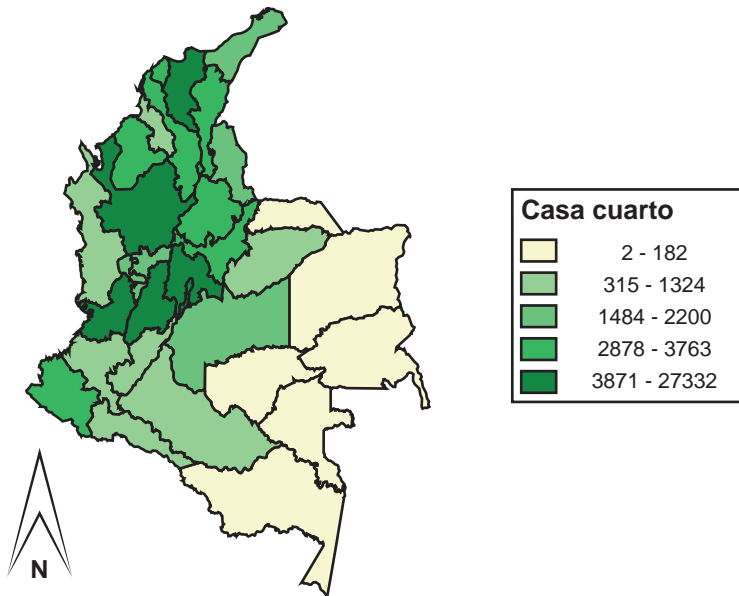
Mapa 16. Departamentos. Número de PM según tipo de vivienda (casa indígena) - 2005.



Fuente: DANE Censo General 2005.

origen, en contraste con Antioquia y Santander, que tuvieron mayor retención de su población. Por su parte, Cundinamarca observó una retención de su población asociada posiblemente con la cercanía a Bogotá y con la baja industrialización y modernización del territorio que no provocó significativos cambios residenciales en su población, lo que conservó una distribución de predominio rural. Así mismo se observaron cambios habitacionales, que reflejan la progresiva densificación de los territorios, pasando de la casa como la forma residencial más habitual a los edificios de apartamentos que responden a nuevas distribuciones económicas y socio-espaciales.


Mapa 17. Departamentos. Número de PM según tipo de vivienda (cuarto) - 2005.



Fuente: DANE Censo General 2005.

Bibliografía

- Arango, G., Peláez, P. y Wolf, G. (1995). *Los cambios en la vivienda en Colombia. Discursos y percepciones*. Medellín: Universidad Nacional de Colombia.
- AIBR. *Revista de Antropología Iberoamericana*. Enero-mayo 2009. 4(1), 142-158. Madrid: Antropólogos Iberoamericanos en Red. Recuperado de www.aibr.org
- Agnew, J. (2008). Spatiality and Territoriality in Contemporary Social Science. *Geopolíticas: espacios de poder y poder de los espacios*. Carlo Emilio Piazzini S. y Vladimir Montoya A. Editores. La Carreta Editores E.U.
- El Colegio de México. Disponible en: <http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=31211538003>
- Fals Borda, O. (1996). *Región e historia. Elementos sobre ordenamiento y equilibrio regional en Colombia*. Bogotá, D. C.: TM Editores.
- Flórez, C (2000). *Las transformaciones sociodemográficas en Colombia, durante el siglo XX*. Bogotá, D. C.: TM Editores.
- Hernández Borge, J. (2006). *Vejez y territorio en Galicia*. Universidad de Santiago de Compostela SEMATA, Ciencias Sociales e Humanidades. Vol. 18, 143-166. Recuperado de http://minerva.usc.es/bitstream/10347/4559/1/pg_143-166_semata18.pdf
- Latour, Bruno (2008). *Reensamblar lo social: una introducción a la teoría del actor-red*. Ediciones Manantial.
- López Ramírez, A. (2008). Migración, remesas y arreglos residenciales de los adultos mayores en México. *Estudios Demográficos y Urbanos*, 23(3), septiembre-diciembre, 513-541.
- Mejía, G. (1999). *Los años del cambio. Historia urbana de Bogotá (1820-1910)*. Bogotá. D. C.: Centro Editorial Javeriano.
- Montañez, G. y Delgado, O. (1999). Espacio, Territorio y Región: conceptos básicos para un proyecto nacional. *Revista Cuadernos de Geografía*, 1(2), 120-134. Bogotá, D. C.: Universidad Nacional de Colombia.
- Puga, D. (2004). El comportamiento residencial de los mayores. Análisis biográfico de la movilidad en la vejez. *Reis. Revista Española de Investigaciones Sociológicas*, 105, 79-102. Centro de Investigaciones Sociológicas, Madrid, España. Disponible en: <http://redalyc.uaemex.mx/src/inicio/ArtPdfRed.jsp?iCve=99717671003>
- San Román, T. (1992). Espacio y ancianidad. *Historia y fuente oral*, núm. 7. Disponible en: *Anal-fabetismo y Política*, pp. 175-178. Publicado por Historia, Antropología y Fuentes Orales. URL artículo Estable:<http://www.jstor.org/stable/27753357>
- Sánchez, D. (2009). Contexto ambiental y experiencia espacial de envejecer en el lugar: el caso de Granada. Universidad Autónoma de Nuevo León. *Papeles de Población*, 60, abril-junio.
- Zambrano, F. y Bernard, O. (1993). *Ciudad y territorio. El proceso de poblamiento en Colombia*. Colombia: Tercer Mundo Editores.



**Los años
de esperanza
de vida perdidos
en adultos mayores.
Bogotá, Cali
y Medellín (1999-2009)**

**Óscar Augusto Castellanos Ospina
Doris Cardona Arango**

Los años de esperanza de vida perdidos en adultos mayores. Bogotá, Cali y Medellín (1999-2009)

Óscar Augusto Castellanos Ospina: Profesional en Mercadeo y Publicidad, magíster en Demografía de la Universidad Nacional de Córdoba – Argentina. Profesional de la Dirección de Censos y Demografía del DANE. Correo electrónico: oacastellanos@dane.gov.co

Doris Cardona Arango: Administradora de Empresas, magíster en Salud Pública y en Epidemiología de la Universidad de Antioquia. Doctora en Demografía de la Universidad Nacional de Córdoba – Argentina. Docente de la Facultad de Medicina de la Universidad CES. Correo electrónico: dcardona@ces.edu.co

Fecha de recepción: 14 de mayo de 2013
Fecha de aceptación: 11 de octubre de 2013

Resumen: en los últimos tiempos, el mundo ha presenciado un desaceleramiento en el crecimiento poblacional que ha modificado su estructura etaria, generando un aumento del número de efectivos en edades mayores y disminuyendo los menores de 15 años; esto ha sido llamado envejecimiento demográfico. El envejecimiento de la estructura por edad y sexo de la población conlleva a que la morbilidad y la mortalidad tengan cambios, pasando de una alta prevalencia en enfermedades transmisibles y parasitarias a una alta incidencia de enfermedades crónicas y degenerativas sobre este grupo de personas en aumento.

El objetivo propuesto fue analizar el perfil de la mortalidad de los adultos mayores en las tres principales capitales de Colombia: Bogotá, Cali y Medellín, entre los años 1999 y 2009, según la lista 6/67 propuesta por la Organización Panamericana de la Salud, comparando su evolución de acuerdo con los años de esperanza de vida perdidos.

Para esto, se usaron los registros de defunción y las proyecciones de población de las tres ciudades, obtenidos del Departamento Administrativo Nacional de Estadística de Colombia (DANE). En el procesamiento de datos se calcularon los indicadores: esperanza de vida al nacimiento, años de esperanza de vida perdidos y años de esperanza de vida perdidos según grupo de causas de mortalidad, usando las rutinas LTPOPDTH, AEVP-6085 y DT0805 Population Analysis Spreadsheets (PAS) del U.S. Bureau of the Census.

Este comparativo permitió reflexionar sobre la necesidad de contar con información propia y de calidad para cada ciudad, que aporte al direccionamiento de los programas de salud encaminados a reducir la incidencia de la mortalidad, el control de enfermedades y la demanda de servicios, los cuales afectan directamente la calidad de vida y el desarrollo social de los adultos mayores.

Palabras clave: mortalidad adulta, años de esperanza de vida perdidos, envejecimiento poblacional.

Abstract: *the whole world has witnessed a rapid population growth, with an increase in life expectancy at birth, leading to a change in age structure, producing the phenomenon called demographic aging population, which is the product of socio-economic, technological, scientific and political transformation of the population. Aging of the structure by age and sex of the population, leads to modifications of the causes of illness and death going from a high prevalence of communicable and parasitic diseases to a high incidence of chronic and degenerative diseases.*

In order to analyze this situation, this work describes the mortality profile by clustering the causes according to the proposal 6/67 of the Pan American Health Organization, comparing their evolution according to the expected life-years lost, in the three main cities of Colombia: Bogotá, Cali and Medellín between 1999 and 2009.

For this purpose, we used death records and population projections, obtained from the National Bureau of Statistics of Colombia (DANE) of the three cities. In data processing three PAS (Population Analysis Spreadsheets) routines were used from the U.S. Bureau of the Census: the LTPOPDTH to perform life tables; the AEVP-6085 to obtain the expected life-years lost by causes and finally the DT0805 to know the contribution of the causes to the life expectancy at birth.

This comparison allowed to consider the need for city-specific information that contribute to addressing health programs, aimed at reducing the incidence of mortality, disease control and the demand for services, which directly affect the quality of life and social development of older adults.

Keywords: *adult mortality, expected life-years lost, demographic aging population.*

1. Introducción

Junto al cambio absoluto por el aumento en sus efectivos, la población mundial avanza hacia el envejecimiento de su estructura por edades, con disminución de los grupos más jóvenes y aumento de la población de adultos mayores (Salgado de Snyder y Wong, 2007). Este aumento hará que para mediados de siglo se igualen algunos grupos poblacionales a nivel mundial (Naciones Unidas, 2002).

El envejecimiento suele describirse sintéticamente como el incremento sostenido de la proporción de personas de 60 y más años con respecto a la población total, lo que resulta de una progresiva alteración del perfil de la estructura por edades, cuyos rasgos piramidales se van transformando a formas rectangulares y, posteriormente, a la inversión de su forma inicial (con una cúspide más ancha que su base). Esta secuencia —a diferen-

cia de lo que ocurre con el envejecimiento de los individuos— no sigue un ritmo cronológico único sino que obedece al sentido, fuerza y persistencia de los cambios de las variables demográficas fundamentales: mortalidad, fecundidad y migración (Villa y Rivadeneira, 2000).

Como lo señala Del Popolo (2001), frente al mundo y América Latina, Colombia vive un proceso de plena y rápida transición demográfica con una natalidad descendente y una mortalidad moderada o baja; transición entendida de acuerdo con la expresión de Vallin (1994) como un proceso que parte de un equilibrio en el crecimiento de la población con una alta mortalidad y fecundidad que se compensan entre sí y que culmina también en equilibrio pero con niveles de fecundidad y mortalidad muy bajos.

Colombia cuenta con 46.941.059 habitantes en el 2013, según la fuente oficial para el país.¹ De acuerdo con las proyecciones de la misma entidad (DANE, 2011), en

1985 el grupo de edades adultas mayores de 60 años y más era de 2.142.219 (6,5% de la población total); en 2013 es de 4.964.793 (10,5% de la población total), y en el año 2025 será de 6.440.000 (12,6% de la población total). Así, este grupo de edad se duplica en 40 años, lo que indica el acelerado crecimiento de este grupo de edad.

Tomar conciencia de esta tendencia exige readecuar las instituciones económicas y sociales para atender demandas hasta hace poco desconocidas. Siguiendo a Villanueva (2000), la transición demográfica y la naturaleza de los problemas de salud de Colombia y de los países en desarrollo económico plantean retos importantes al sistema de salud, debido al incremento y muerte por enfermedades crónicas y degenerativas que requieren servicios especializados, de alto costo y atención multidisciplinaria, los cuales evidencian diferencias en el acceso y en la asistencia médica de alta calidad entre las distintas clases socioeconómicas.

¹ Ver reloj poblacional de Colombia en http://www.dane.gov.co/reloj/reloj_animado.php consultado el 28 de febrero de 2013.

El índice de Años de Esperanza de Vida Perdidos (AEVP) propuesto por el doctor Eduardo Arriaga (1997) es útil para sintetizar el nivel y cambio de la mortalidad entre poblaciones y en distintos momentos del tiempo al eliminar los efectos propios de la estructura por edad y sexo que se hacen presentes en los indicadores tradicionales como las tasas brutas, tasas específicas por edad y en la estandarización de las mismas tasas. La esperanza de vida al nacimiento permite observar adecuadamente el nivel de la mortalidad pero no permite evidenciar el cambio de la mortalidad, toda vez que indica los años de vida que una población vive como consecuencia de los niveles de mortalidad. «Un cambio en la esperanza de vida es un cambio en los años de vida por el cambio en la mortalidad y no directamente el cambio de la mortalidad» (Arriaga, 2011).

Aunque no es del todo perfecto, el autor al explicar el indicador resalta que «Cuando el análisis se hace por causas de muerte, esta metodología ofrece, al menos, una ventaja importante, como es el hecho de medir con el mismo índice el nivel y cambio en la mortalidad en forma tal que refleje cercanamente el cambio en el nivel general de mortalidad» (Arriaga 1997). Continúa diciendo que «Con base en las tasas específicas de mortalidad por edades, se calcula la vida promedio que la población vive (esperanza de vida) y se determina cuántos años de vida se pierden considerando la hipótesis sobre cuántos años debería vivir». Por lo que el indicador busca establecer cuántos años de vida en

promedio pierde una población por la muerte de las personas a determinadas edades.

En Bogotá, Cali y Medellín se encuentra el mayor desarrollo demográfico y económico del país. Siendo las principales ciudades de Colombia, presentan diferencias sociales, económicas, demográficas, geográficas y culturales entre sí, pero unidas conforman lo que se denomina coloquialmente el triángulo de oro de Colombia. Por este motivo, se quiso contribuir al análisis demográfico del envejecimiento de estas tres principales ciudades de Colombia, comprendiendo el reto que enfrenta la sociedad colombiana ante la mortalidad en personas entre 60 y 85 años, relacionando las causas de muerte con el cambio en las esperanzas de vida.

En la primera parte del artículo se describen los materiales y métodos utilizados para la construcción del indicador, seguidos de los resultados obtenidos divididos en cuatro apartados: la diferencia de años de esperanza de vida perdidos, años de esperanza de vida perdidos según grupo de causas de mortalidad, incremento en los años de esperanza de vida perdidos según grupo de causas y un análisis desagregado por causas seleccionadas según mayor índice de AEVP. Se aporta una discusión contextualizada de acuerdo a lo que ocurría en Latinoamérica a finales del siglo pasado, y por último las conclusiones.

2. Materiales y métodos

Como estudio descriptivo y transversal en el tiempo,² se utilizan los certificados de defunción de las Estadísticas Vitales para personas con 60 años y más de edad fallecidas en las ciudades de Bogotá, Medellín y Cali en los trienios 1999-2001 y 2007-2009³ reportados al Depar-

tamento Administrativo Nacional de Estadística (DANE) como fuente de información secundaria (Colombia y Departamento Administrativo Nacional de Estadística, 2011). Para los denominadores se usan las proyecciones de población con base en el censo 2005 (Colombia y Departamento Administrativo Nacional de Estadística, 2012).

El primer índice calculado fue la esperanza de vida al nacimiento (e^0), la cual se obtiene generalmente de tablas de vida como resumen del impacto de la mortalidad en todas las edades para ambos sexos y no está afectada por la estructura etaria de la población, permitiendo observar la cantidad de años que una cohorte hipotética de nacimientos (por lo general 100.000) alcanza a vivir bajo el supuesto de una mortalidad constante en todos los grupos etarios hasta que se extingue.

Se utiliza la esperanza de vida temporaria⁴ entre 60 y 85+ años de edad para medir adecuadamente la esperanza de vida para el grupo de los adultos mayores, eliminando el efecto de la mortalidad de otros grupos etarios menores. Se agrupan las edades mayores a 85 años para eliminar las fluctuaciones que se producen en estas edades por errores en el dato como la mala declaración de la edad.

Como se explicó en la introducción de este trabajo, el índice AEVP⁵ fue formulado por Eduardo Arriaga para medir el nivel y el cambio de mortalidad por causas de muerte y edad, sin las limitaciones intrínsecas de la tasa bruta, la específica por edad de la mortalidad, la estandarización de las mismas y la esperanza de vida.⁶ Arriaga, al considerar las falencias del indicador (distinto al cálculo de la esperanza de vida), manifestó que se requiere limitar las edades bajo estudio y dejar fuera del análisis las edades de un grupo de edad abierto, debido a que la declara-

² Al no disponer de datos longitudinales para las cohortes analizadas, el estudio dispone de información transversal por año de las Estadísticas Vitales y las proyecciones de población del DANE.

³ Se obtiene un promedio móvil trianual con el fin de corregir las fluctuaciones anuales en los datos de las defunciones, y como denominador la población proyectada para el año de la mitad del trienio. El promedio entre los dos periodos permitió observar en resumen lo que sucedió en el decenio analizado con el indicador.

⁴ $e_x = (T_x - T_{x+n}) / 1$, donde (e) es la esperanza de vida del intervalo de edades que se analiza, T_x es la población total de la tabla de vida y 1, son los sobrevivientes de una edad específica en la tabla de vida.

⁵ $AEVP_x = 1 - e_x$, donde (1) es el número de años del intervalo de edades que se analiza y (e) es la esperanza de vida del intervalo de edades.

ción de las causas de muerte en edades avanzadas no es confiable, por lo que limitar la edad superior se traduce en una ventaja (Arriaga, 1997); adicionalmente, porque se necesita saber cuál es el máximo número de años que una población puede vivir y para el grupo abierto de edad es indeterminado. Por lo anterior, para el caso de estudio, el intervalo de edades se definió entre 60 y 84 años de edad.

El indicador años de esperanza de vida perdidos se deduce de tablas de mortalidad calculadas por los autores con la planilla LTPOPDTH (dado que no existen tablas de vida oficiales para las ciudades y años analizados) y de la descomposición de la esperanza de vida a través de la utilización de dos rutinas de análisis poblacional, PAS (del inglés: Population Analysis Spreadsheets, PAS), del U.S. Bureau of the Census: LTPOPDTH y AEVP6085.

Primero, se utiliza la rutina PAS titulada DT0805 para conocer el nivel general de la mortalidad por la diferencia de la esperanza de vida desde el nacimiento a la edad abierta de 80 y más años por edades quinquenales en las tres ciudades estudiadas. Segundo, se aplica la rutina titulada «LTPOPDTH» (del inglés Life Table Population Death), que sirve para construir tablas de vida a partir de datos de población proyectada y defunciones por edad de las Estadísticas Vitales; se obtuvieron para cada sexo, ciudad y trienio analizado. De estas tablas de vida o mortalidad, se extraen las funciones insumo para la obtención del indicador AEVP como la esperanza de vida (e^x), los sobrevivientes (l_x) y la población (L_x). La última rutina PAS utilizada fue la titulada

⁶ Se considera que la tasa bruta de mortalidad está afectada por la estructura por edades de la población, y que la tasa específica por edad no permite comparar el nivel de mortalidad entre poblaciones que tienen niveles de mortalidad parecidos en sus edades. «La estandarización ofrece solo rangos de los niveles de las tasas estandarizadas, pero no necesariamente su nivel de mortalidad» (Arriaga, 2011) y la esperanza de vida mide adecuadamente el nivel pero no permite medir adecuadamente el cambio relativo de la mortalidad. (Arriaga, 1997).

⁷ Eduardo Arriaga (2011), citando a González Galé (1951), indica que es aquella perteneciente al periodo de vida entre dos edades específicas. Es decir, la esperanza de vida desde la edad x hasta la edad de $x+n$. En símbolos: $e_x = (T_x - T_{x+n}) / l_x$ (Arriaga, 2011).

AEVP6085 que permite obtener las esperanzas de vida temporarias⁷ del grupo de edad 60-85+ años y conseguir los años de esperanza de vida perdidos para el grupo de edad 60-84 años para las tres ciudades, discriminados por sexo y periodo de referencia, permitiendo conocer la dinámica del nivel y cambio de la mortalidad durante el decenio analizado.

El índice AEVP se calcula sobre las categorías de causas de defunción

agrupadas por el DANE según la lista OPS 6/67 de la Organización Panamericana de la Salud, la cual es un ordenamiento a seis grandes grupos y 67 subcategorías de la décima clasificación internacional de enfermedades y problemas relacionados con la salud (CIE-10), que facilita su comparabilidad nacional o internacional. Los grandes grupos de causas se dividen en las siguientes categorías:

Tabla 1. Grandes grupos de causas. Lista 6/67 de la OPS (CIE-10).

100. Enfermedades infecciosas
200. Neoplasias (tumores)
300. Enfermedades del sistema circulatorio
500. Causas externas
600. Todas las demás causas
700. Síntomas, signos y afecciones mal definidas

Fuente: Elaboración propia a partir de grandes grupos de causas lista 6/67 OPS.

Seguidamente y de acuerdo con los niveles de AEVP obtenidos para estos siete grandes grupos de causas, se seleccionan y desagregan las subcategorías de causas de los tres principales grandes grupos de

causas que presentaron los mayores niveles de AEVP para cada ciudad y por sexo, dejando fuera de la selección aquellas subcategorías que estuvieran por debajo de 0,5 AEVP (ver figuras 6 y 7).

3. Resultados

De acuerdo con las Estadísticas Vitales oficiales, la ciudad de Bogotá Distrito Capital registró el mayor número de defunciones en los dos periodos de estudio, seguida de Medellín (capital del departamen-

to de Antioquia) y Cali (capital del departamento del Valle del Cauca). Situación apenas lógica dado que el número de habitantes es mayor en la capital del país y probablemente tiene una estructura poblacional más envejecida, como se presenta en la siguiente tabla:

Tabla 2. Promedio de defunciones por trienio y ciudad para ambos sexos. Edad 60-85+.

Trienio	Ciudad	Defunciones	Población
1999-2001	Bogotá	19.514	6.302.881
	Medellín	8.875	2.065.871
	Cali	8.793	1.998.887
2007-2009	Bogotá	24.738	7.155.052
	Medellín	10.990	2.291.378
	Cali	10.208	2.194.753

Fuente: DANE-Estadísticas Vitales. Registros de defunciones Bogotá, Medellín y Cali. 1999-2009 y cálculos del autor.

Hay que considerar que el nivel de omisión o no reporte de la información de las estadísticas vitales de defunciones en cada ciudad es desconocido oficialmente, lo que requiere de evaluaciones a la calidad, oportunidad y cobertura de la información de las Estadísticas Vitales para definirlo. Una investigación DANE-UNFPA (2011) inédita realizó un estudio de cabalidad para el periodo comprendido entre 2006 y 2008 para Bogotá, Antioquia y Valle del Cauca, indicando una no cobertura en defunciones de 12,8; 14 y 16,1 respectivamente, por lo que los incrementos pueden tener implícito un mejoramiento en la cobertura entre los periodos analizados, y por lo tanto, las proyecciones de población pueden tener varios procesos de suavizamiento para regular las fluctuaciones que

se presentaron en el registro. Sin embargo, los registros son aceptables para realizar estudios que se acerquen a la situación de mortalidad de cada ciudad.

4. Años de esperanza de vida perdidos

En los hombres, la ciudad que registró una mayor esperanza de vida (o años en promedio que se espera que una población viva, de mantenerse las condiciones de mortalidad observadas para ese periodo de tiempo durante toda su vida) en el trienio 1999-2001, fue Bogotá con 73 años, seguida de Medellín y Cali que se mantuvieron en 68 años de esperanza de vida al nacer (tabla 3).

el segundo trienio 2007-2009, el aumento fue de un año, cuyas diferencias entre las ciudades se observaron de manera decimal.

La diferencia entre hombres y mujeres marcó una pauta mayor en Medellín con 10,04 años, y le siguen Cali con 9,11 y Bogotá con 6,02 años, lo que muestra una gran diferencia en el nivel de mortalidad diferencial por sexo de las tres ciudades, siendo el más equitativo entre los sexos, Bogotá.

La esperanza de vida permitió observar el nivel de la mortalidad en las tres ciudades, al conocer el número promedio de años de vida que viven bajo las condiciones de mortalidad observadas al inicio y final del decenio y los cambios en los mismos. Sin embargo, para medir el cambio de la mortalidad en los adultos mayores de acuerdo con los años de esperanza de vida que se pierde, se hizo necesario primero obtener la esperanza de vida temporaria entre las edades de 60 y 85+ para poder eliminar el efecto que produce la mortalidad de otros grupos de edad en la esperanza de vida adulta. Se utilizó la rutina PAS llamada AEVP6085, que comparó la esperanza de vida temporaria para el grupo de edad quinquenal de 60 a 85+ años para ambos sexos, ciudad y periodo de análisis; posteriormente, esta rutina permitió distribuir las defunciones por causas de grupos quinquenales de edad, sexo, ciudad y trienio, para finalmente obtener los Años de Esperanza de Vida Perdidos (AEVP) entre las edades quinquenales de 60 a 84 años por causas de muerte.

Sobre la hipótesis de que los adultos mayores deberían vivir hasta los 85 años en promedio, según la tabla 4, en el primer trienio para los hombres Cali fue el que mayor número de Años de Esperanza de Vida perdió del grupo con 7,11 años, seguido de Bogotá y Mede-

Tabla 3. Esperanza de vida al nacimiento, según trienio y sexo. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.

Sexo	Trienio	Ciudad		
		Bogotá	Medellín	Cali
Hombres	1999-2001	73,32	68,02	68,81
	2007-2009	76,24	73,77	71,78
	Diferencia	2,92	5,74	2,97
Mujeres	1999-2001	80,16	80,19	78,85
	2007-2009	81,43	81,67	79,97
	Diferencia	1,27	1,48	1,12
Diferencia H-M	2007-2009			
	1999-2001	6,02	10,04	9,11

Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (DT0805). Diferencias tomando como base el periodo final (2007-2009).

Los hombres de Bogotá y Cali mantuvieron una diferencia similar de 2,9 años de incremento en sus esperanzas de vida de un trienio a otro; en tanto que Medellín logró un significativo avance en el aumento de sus años de esperanza de vida duplicando a Bogotá y Cali en su diferencial con 5,74 años de vida en el decenio.

Entre las causas que se atribuyen al aumento en la esperanza de vida, están las innovaciones tecnológicas, la mejora sanitaria y de

la higiene con control de enfermedades infecciosas y parasitarias, mejor alimentación y eficientes programas de atención primaria en salud pública y servicios en salud, siendo el acceso a los mismos lo que genera las principales diferencias entre las poblaciones y entre los sexos.

En las mujeres la mayor esperanza de vida al nacer en el primer trienio la compartió Medellín con Bogotá en 80 años, mientras que Cali se aproximaba a los 79 años. Para

llín con 6,65 y 6,46 AEVP respectivamente. Para el segundo trienio, todas las ciudades presentan una reducción en los Años de Esperanza de Vida que se pierde por causa de la mortalidad en los hombres, teniendo como tope los 85 años de esperanza de vida.

Bogotá fue la ciudad que logró reducir más el indicador AEVP de las tres ciudades al finalizar el decenio con 5,75 Años de Esperanza de Vida Perdidos, aunque fue muy similar a Medellín, que lo ubicó en 5,80. Cali es el más rezagado con 6,61 AEVP, lo que se reflejó en la diferencia absoluta entre un trienio y el otro, donde Bogotá redujo el indicador en 0,90 AEVP seguido por Medellín con 0,66 y Cali con 0,50 AEVP.

En cuanto a las mujeres, al igual que en los hombres, sigue siendo Cali el que tuvo un mayor número de Años de Vida Perdidos al inicio del decenio con 5,24; le sigue Bogotá con 4,91 y Medellín con 4,83 AEVP (tabla 4). Al final del primer decenio del milenio, la ciudad con el menor indicador fue Medellín con 4,09 AEVP y el que más redujo la diferencia durante el decenio analizado con 0,74 AEVP, aunque las diferencias son centesimales con las otras ciudades.

La diferencia se amplió un poco más entre sexos durante todo el periodo de referencia. En promedio, Cali es la ciudad con un mayor rango de diferencia entre sexos con 1,93 AEVP, mientras que Bogotá y Medellín tuvieron indicadores muy similares con 1,65 y 1,67 AEVP de diferencia, a pesar de que Medellín fue el que mayor diferencia de esperanza de vida al nacimiento presentó entre hombres y mujeres con 10,4 años en promedio durante el decenio (tabla 3), observándose cómo con este procedimiento y al usar esperanzas de vida temporarias, la medición del nivel recibe otra perspectiva.

La poca reducción del indicador durante el decenio evidencia la velocidad para observar cambios significativos en el indicador para los adultos mayores, sujeto al avance en la intervención para el mejoramiento de la calidad de vida, con políticas que van desde el mejoramiento de la calidad en la información de Estadísticas Vitales, hasta el acceso de programas y planes de salud para este grupo poblacional.

5. Años de esperanza de vida perdidos según grupo de causas de mortalidad

Al calcular los AEVP por grandes grupos de causas, según la lista 6/67 de la OPS para los hombres, se pudo identificar como el grupo principal de causas de muerte en las tres ciudades a las enfermedades del sistema circulatorio, seguido de las demás causas y las neoplasias o tumores, lo que las evidencia como los grupos de causas de mortalidad propias de poblaciones envejecidas.

Los siguientes grandes grupos de causas están por debajo de 0,5 AEVP como las causas externas y las enfermedades infecciosas y las causas mal definidas referidas a la

calidad de la certificación (gráficos 1 y 2). Lo anterior pone de muestra el reto social para los gobiernos, una vez cambiada la tendencia del perfil de mortalidad por el envejecimiento poblacional, para garantizar los servicios de salud que demanda este creciente grupo de personas potencialmente dependientes.

De acuerdo con el gráfico 2, para las mujeres varió levemente el orden de los tres grupos de causas identificados igualmente como los principales entre los hombres: las enfermedades del sistema circulatorio son el principal grupo de causas para las tres ciudades de acuerdo con el mayor número de AEVP que presentó. El segundo lugar difiere levemente entre las tres ciudades, pues para las mujeres bogotanas son las neoplasias, mientras que para las mujeres de Medellín y Cali es el grupo de todas las demás causas (que incluye la diabetes mellitus), presentándose en forma contraria para el tercer grupo con mayor índice de AEVP entre las ciudades.

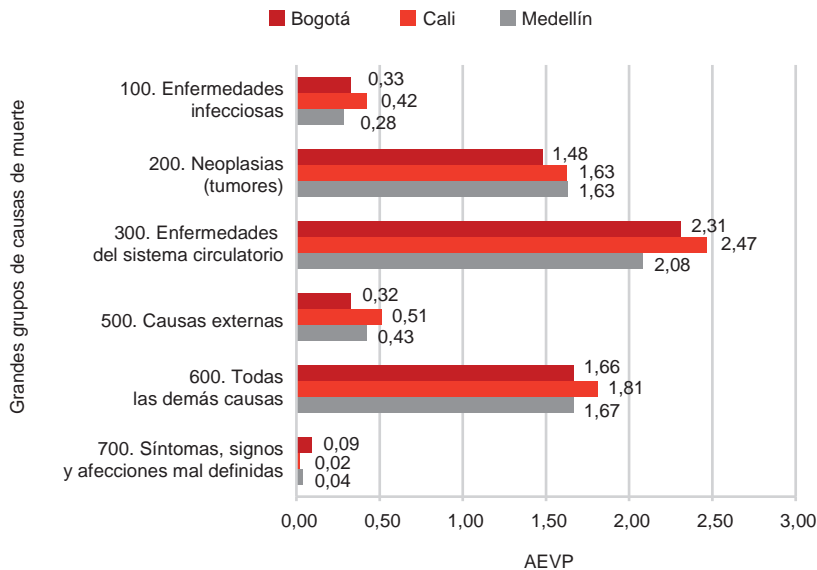
Se destaca la oportunidad que puede representar el apoyo intermunicipal de estas tres ciudades en su objetivo común por reducir los indicadores de AEVP. Por ejemplo, si Cali y Medellín obtuvieran el apoyo interinstitucional de Bogotá (su experiencia) para reducir el indicador

Tabla 4. Diferencia de los años de vida perdidos, según trienio y sexo, para los años 60-85. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.

Sexo	Trienio	Ciudad		
		Bogotá	Medellín	Cali
Hombres	1999-2001	6.65	6.46	7.11
	2007-2009	5.75	5.80	6.61
	Diferencia	0.90	0.66	0.50
Mujeres	1999-2001	4.91	4.83	5.24
	2007-2009	4.19	4.09	4.63
	Diferencia	0.72	0.74	0.61
H-M	1999-2001	1.65	1.67	1.93
	2007-2009			

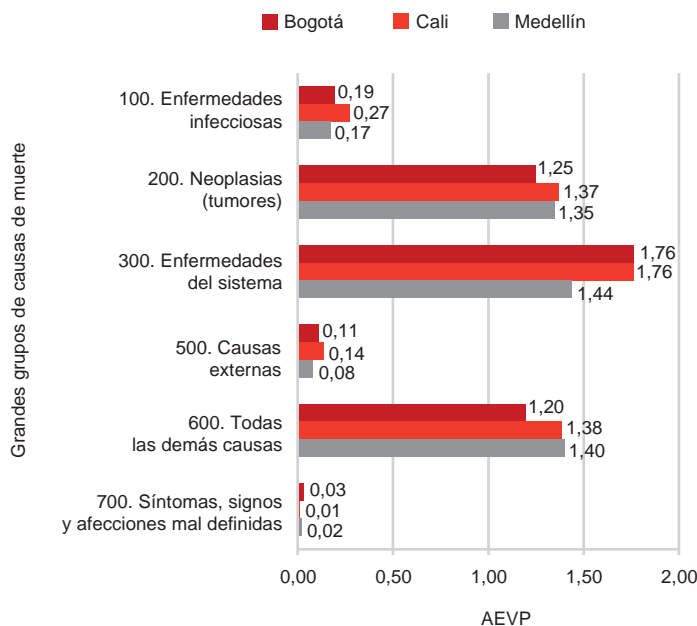
Fuente: DANE. Registros de defunciones y proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6585). Diferencias tomando como base el periodo inicial (1999-2001).

Gráfico 1. Años de esperanza de vida perdidos en los hombres desde la edad de 60 a 85 años, por grandes grupos de causas, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

Gráfico 2. Años de esperanza de vida perdidos en las mujeres desde la edad de 60 a 85 años, por grandes grupos de causas, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

de AEVP por neoplasias en mujeres, y lograr como mínimo el indicador de Bogotá (gráfico 2), reducirían en 0,12 y 0,15 los AEVP, lo que

se traduciría en años de esperanza de vida ganados por esta causa. De igual forma, si Bogotá y Cali obtuvieran el apoyo de Medellín para

la reducción del indicador AEVP por enfermedades del sistema circulatorio en hombres, obteniendo como mínimo el indicador de Medellín reducirían en 0,39 y 0,23 los AEVP de cada ciudad (gráfico 1) y en lo sucesivo con cada categoría, como una acción en la lucha por la reducción de las enfermedades y mortalidad de los adultos mayores.

Es importante resaltar la calidad de la certificación de causas, la cual requiere que los médicos certificadores del hecho vital diligencien adecuadamente las causas que conllevan al fallecimiento de los adultos mayores, sin omitir causas de base para una adecuada codificación del perfil de mortalidad.

6. Reducción en Años de Esperanza de Vida Perdidos según grandes grupos de causas

La reducción o aumento del indicador en términos absolutos permitió reflejar el impacto de las decisiones en materia de salud para los adultos mayores de las tres ciudades, analizadas durante el periodo de referencia 1999-2009, como consecuencia de los factores relacionados, más arriba, del aumento en la esperanza de vida.

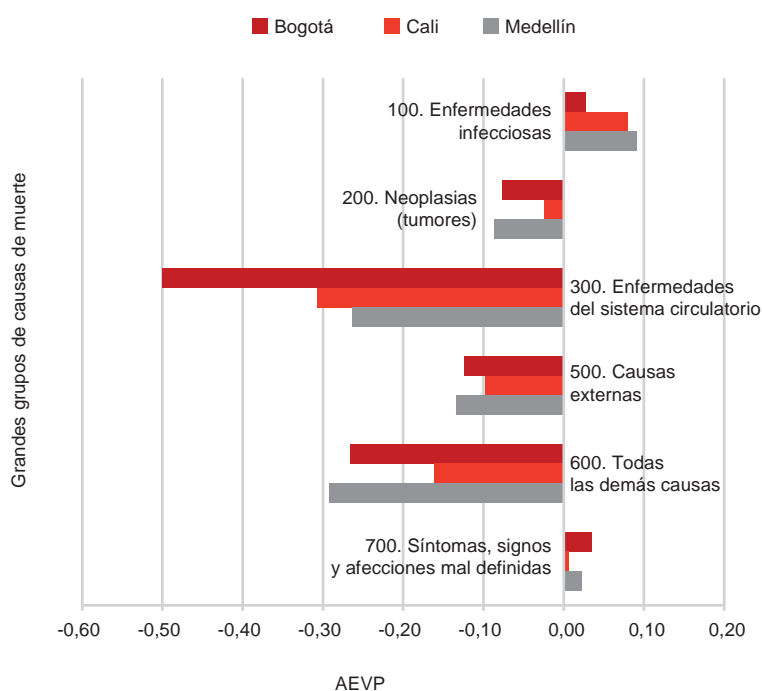
Para los hombres de Bogotá y Cali, la mayor reducción de AEVP en el decenio fue debido a las enfermedades del sistema circulatorio, que a su vez representó un aumento en la esperanza de vida temporaria de 0,50 y 0,31 años como se observa en el gráfico 3. En cuanto a los hombres de Medellín, el grupo de causas con mayor reducción de AEVP se dio en el grupo de todas las demás causas (que incluye la diabetes mellitus) con una reducción de 0,29 AEVP. Lo anterior en contraste con un aumento de

AEVP del grupo de enfermedades infecciosas que presentó una mayor proporción en Medellín que en Cali y Bogotá, y un aumento en la categoría síntomas, signos y afecciones mal definidas, lo que a su vez representó un incremento en el indicador AEVP (gráfico 3).

Observar la disminución de AEVP en estos grupos de causas de mortalidad da luces sobre las decisio-

nes de los gobernantes en el mejoramiento de las condiciones de vida para los adultos mayores y el acceso a los servicios y programas en salud durante el decenio analizado, esfuerzos que deben continuar sin descuidar los otros grupos de enfermedad como las enfermedades infecciosas y la calidad en la declaración de causas al momento de la certificación.

Gráfico 3. Incremento y reducción de los AEVP en los hombres desde la edad 60 hasta 85 años por grandes causas de muerte, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

Para las mujeres, se observó una situación similar que en los hombres: Cali compartió la situación de Bogotá con una mayor reducción de la categoría de enfermedades del sistema circulatorio, lo que se traduce en un aumento de 0,46 y 0,39 años en la esperanza de vida temporaria, respectivamente. Medellín presentó una mayor reducción por todas las demás causas con una ganancia de 0,39 años. También se evidenció un incremento en la categoría de enfermedades

infecciosas (no así para Bogotá) y del grupo de síntomas, signos y afecciones mal definidas, que fue levemente menor que en los hombres. (gráfico 4).

El control de las defunciones por enfermedades infecciosas en las últimas décadas ha sido lo que con más años ha contribuido a la esperanza de vida de las poblaciones humanas, con medidas aplicadas en salud pública como el fomento a la cultura de prevención (desinfección, higiene, cocción de los ali-

mentos en los hogares, entre otros). Sin embargo, se observó un repunte de este grupo de causas para los adultos mayores en las tres ciudades comparadas en el último decenio, lo que prende las alarmas sobre mantener o aumentar el control preventivo y activo de este grupo de causas como las infecciones respiratorias o intestinales, septicemia, meningitis o tuberculosis.

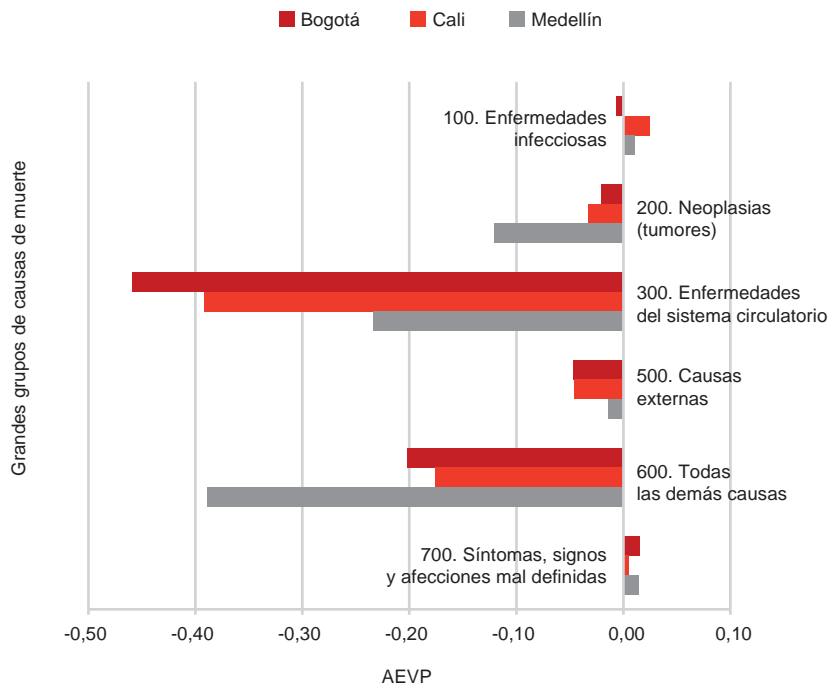
Se marca en las mujeres de Medellín el probable mejoramiento en el acceso a la prevención y atención en salud realizadas para reducir los índices AEVP por las demás causas y enfermedades del sistema circulatorio o la mejor calidad en la certificación de defunción, aumentando en suma 0,62 años la esperanza de vida temporaria. Tanto en hombres como en mujeres, el grupo de causas que más AEVP redujo durante el decenio fueron las enfermedades del sistema circulatorio, lo que se traduce en años de esperanza de vida ganados.

A continuación se compara la contribución de todas las causas de fallecimiento por edades quinquenales entre 60 y 84 años de hombres y mujeres de las tres ciudades: en los hombres, Cali y Medellín presentan una curva de cúspide llana ascendente a la edad inicial de 60-64 años para iniciar su descenso en el siguiente grupo quinquenal; Bogotá continúa con un leve ascenso hasta el grupo de edad de 70-74 años, para empezar a decrecer de manera pronunciada por debajo de los 0,5 AEVP al llegar al grupo de 80-84 años, indicando que en Bogotá el grupo de 65-74 años contribuye con más AEVP por el efecto de toda su mortalidad que en el de Cali y Medellín (gráfico 5).

Para el caso de las mujeres, las tres ciudades realizan un ascenso hasta la edad de 70-74 años para iniciar su descenso pronunciado hasta el grupo de 80-84 años por

debajo de los 0,5 AEVP, siendo el de Bogotá respecto a Medellín mayor el nivel de Cali y muy similar (gráfico 5).

Gráfico 4. Incremento o reducción de los AEVP en las mujeres desde la edad de 60 hasta 85 años por grandes causas de muerte, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

Es de resaltar la modelización que hace cada ciudad tendiente a generar la forma parabólica de Bogotá a medida que la esperanza de vida aumenta en cada grupo de edad y sexo con la reducción de los AEVP.

7. Análisis desagregado de los principales grupos de causas

De acuerdo con los tres principales grandes grupos de causas de defunción observados que más aportan AEVP de los siete agrupamientos de la lista 6/67 de la OPS, se hace necesario profundizar en el análisis, indagando en cada grupo por sexo, con el fin de desagregar las subcategorías que más AEVP aportan al grupo de causas y a la

mortalidad total de los adultos mayores, con el fin de que sean tenidas en cuenta por la Administración Pública en su toma de decisiones correspondientes. Se seleccionaron aquellas que representan más del 50% del impacto en años de esperanza de vida perdidos resumidos en los gráficos 6 y 7 para cada sexo.

Se observó anteriormente, para las tres ciudades, como la principal categoría de causas de muerte, a las enfermedades del sistema circulatorio propias de las poblaciones que comienzan su etapa de envejecimiento, las cuales, siguiendo a Omran (1971), se identifican en la tercera fase de la transición epidemiológica con la aparición de enfermedades degenerativas y producidas por el hombre.

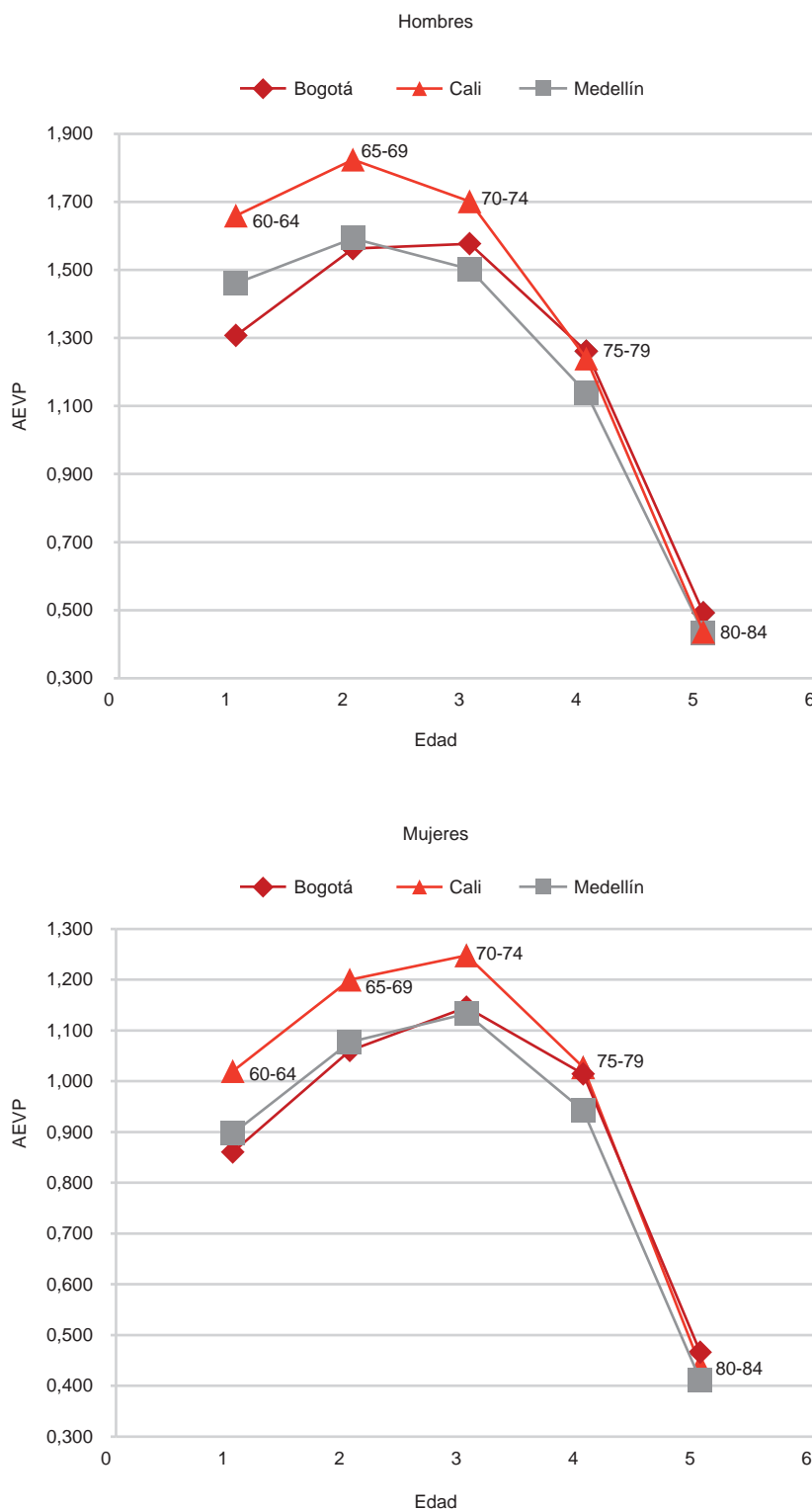
Se destacaron en el mismo grupo de causas, las enfermedades isquémicas del corazón con casi el

50% del peso relativo del grupo; de manera que, si se redujera la mortalidad por isquemias del corazón, cada adulto mayor ganaría 1,20 años de esperanza de vida en promedio para las tres ciudades, entrando probablemente en la siguiente fase de la transición epidemiológica descrita por Omran (1998) como de declinación de mortalidad cardiovascular. Le siguen las enfermedades cerebrovasculares como segunda subcategoría de causa identificada; en este caso, Bogotá aumentaría en 0,49 años su esperanza de vida, Cali aumentaría 0,61 años y Medellín, 0,45 años (gráfico 6).

Las mujeres comparten las mismas dos subcategorías de causas por enfermedades del sistema circulatorio de los hombres, propias de la tercera etapa de la transición epidemiológica, pudiendo aumentar su esperanza de vida temporal en 1,22 años para Bogotá, Cali en 1,29 y Medellín en 1,07 años, de llegarse a reducir en su totalidad la mortalidad por isquemias del corazón y enfermedades cerebrovasculares, lo que requiere una mayor intervención en vigilancia epidemiológica y en salud pública, mejorando la calidad de los registros (gráfico 7).

La segunda categoría identificada corresponde a todas las demás causas, de las cuales son representativas las enfermedades de las vías respiratorias, diabetes mellitus y resto de enfermedades del sistema digestivo, tanto en hombres como en mujeres y con una pequeña variabilidad en AEVP entre las ciudades. Sobresale Medellín como la ciudad con mayor nivel de mortalidad por estas causas (gráficos 6 y 7), probablemente debido a los hábitos de vida y alimentarios, resaltándose las enfermedades de las vías respiratorias inferiores prevalentes en pacientes fumadores e inmunodeprimidos, aunque hay que tener presente el inconvenien-

Gráfico 5. Contribución de todas las causas por edad a los años de esperanza de vida perdidos de los adultos mayores, según sexo. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

te de calidad en los datos como se ha venido expresando.

Para el tercer grupo de neoplasias, sobresale en los hombres el cáncer de estómago, otros tumores y el cáncer de próstata con mayor pérdida de años de vida, observándose un mayor nivel de mortalidad en Bogotá por cáncer de estómago, en Medellín por otros tumores malignos y en Cali por cáncer de próstata, lo que puede corresponder a una falta de prevención y mayores controles de la población adulta.

En las mujeres, la categoría de neoplasias indicó una variabilidad mayor en las subcategorías de causas de las tres ciudades, lo que obligó a una agrupación mayor para obtener más del 50% del impacto de AEVP en conjunto:

En Bogotá, el cáncer de mama junto con el de estómago representaron las dos mayores subcategorías de causas de mortalidad; en Cali fue el tumor maligno de mama, mientras que en Medellín, y con una amplia diferencia, se observa el cáncer de tráquea, bronquios y pulmón, el cual se relaciona al consumo de cigarrillo.

Si Medellín logra controlar el nivel de mortalidad por cáncer de pulmón y reducirlo al nivel de Bogotá, aumentaría en 0,15 su esperanza de vida temporaria para los adultos mayores, mientras que Cali la aumentaría en 0,6 años de esperanza de vida.

8. Discusión

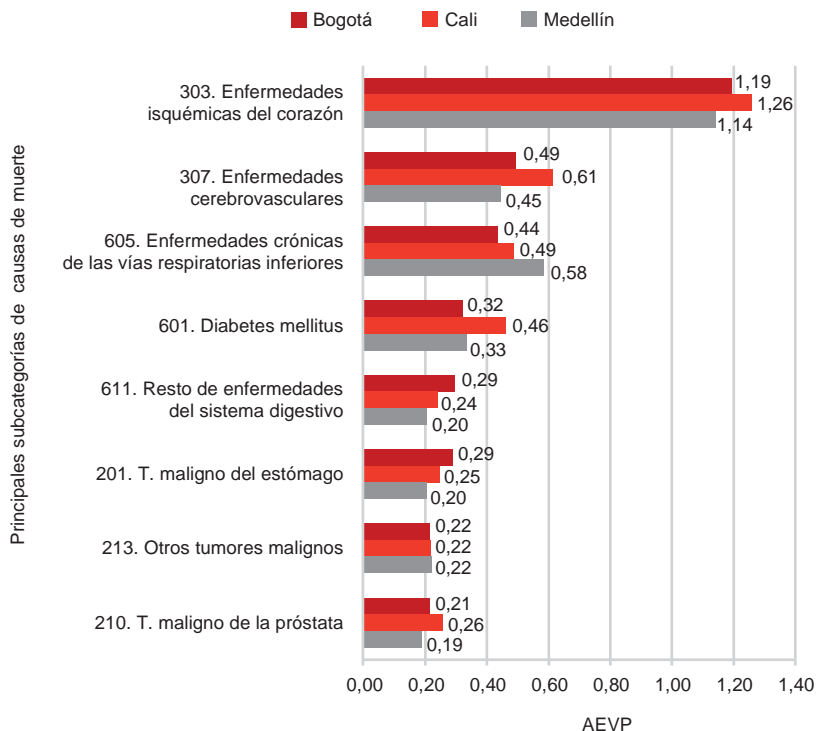
En este apartado, se buscó contextualizar los niveles y cambios de la mortalidad de las tres ciudades con lo que ocurría en Latinoamérica a finales del siglo pasado, de acuerdo con el análisis publicado por CELADE/CEPAL en el 2003, matizado por la desigualdad social

y económica que se refleja en los niveles de mortalidad y morbilidad (González y Ham, 2007). Como resultado, el sistema de salud se ve y seguirá enfrentando un doble problema: aumento en la incidencia de

enfermedades crónico-degenerativas y la persistencia de enfermedades transmisibles, con demanda de prestaciones y tratamientos (Cardona y Peláez, 2010).

mayores. Para las familias, continúan explicando que la demanda se traducirá en transferencias hacia los adultos mayores que no estén cubiertos por algún sistema de protección social o cuya cobertura sea insuficiente para mantener un nivel de vida satisfactorio, situación obligatoria para la población económicamente activa (CELADE/CEPAL et al., 2003).

Gráfico 6. AEVP en los hombres desde la edad 60 hasta 85 años por subcategorías de causas de muerte, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

Según la Organización Mundial de la Salud en su Informe sobre la salud en el mundo (OMS, 2008), «un historial de escaso crecimiento económico es también un historial de estancamiento de los recursos para la salud (...) lo cual dejaba poco margen para ampliar el acceso a la atención sanitaria o transformar los sistemas de salud». Según este organismo, el hecho de no reconocer la necesidad de expertos en temas distintos de las disciplinas sanitarias tradicionales ha condenado al sector de la salud a ver sus sistemas sometidos a un grado insólito de incompetencia e ineficacia que la sociedad no se puede permitir.

Siguiendo el informe de CELADE/CEPAL et al. (2003), el incremento de adultos mayores en las sociedades actuales plantea nuevos retos a las políticas sociales y públicas para atender las necesidades básicas y las demandas en salud de este grupo de la población. Se manifiesta que la ausencia del Estado en los sistemas de protección social actual para este grupo poblacional recaerá sobre los gobiernos y las familias; explicando que para los gobiernos esta demanda se manifestará en la exigencia de extender la cobertura de los sistemas de seguridad social y de aplicar o ampliar los programas de protección social para los adultos

Para las tres ciudades de Colombia, la esperanza de vida al nacer ha aumentado entre 2,9 y 5,7 años en los hombres, y entre 1,1 y 1,5 años para las mujeres durante los períodos estudiados, alcanzando la esperanza de vida de los países más desarrollados (OMS, 2008), sin tener su mismo desarrollo y en un menor lapso de tiempo (González y Ham, 2007), lo que indica una reducción en las defunciones 1999-2009, un descenso en la mortalidad infantil y de la fecundidad, aumentando la población en las cohortes de mayor edad, lo que genera cambios en la transición demográfica de las tres ciudades.

En América Latina y el Caribe, las enfermedades del sistema circulatorio son la característica específica más destacable en el perfil de mortalidad de la población adulta mayor en las últimas dos décadas del siglo XX, y por lo tanto tuvo la contribución más significativa en la ganancia de la esperanza de vida de ambos sexos (CELADE et al., 2003).

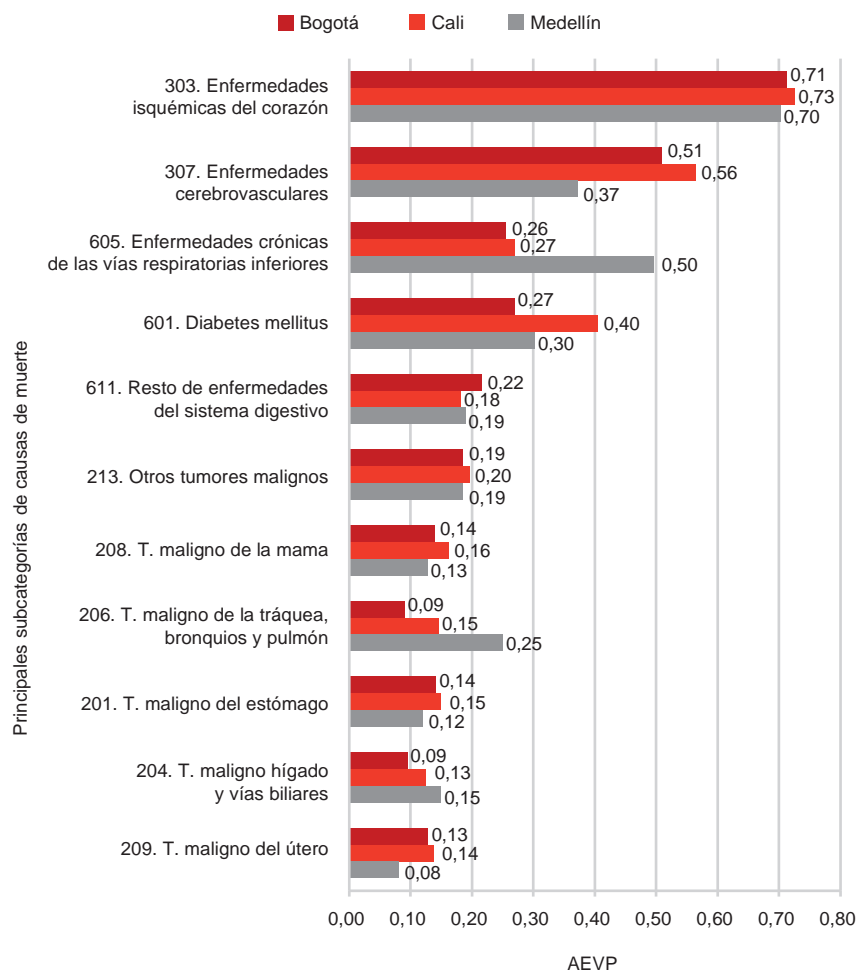
Para las ciudades de Bogotá, Cali y Medellín, pasado el primer decenio del siglo XXI, se pudo observar que las enfermedades del sistema circulatorio son la principal causa de muerte entre los hombres y mujeres adultos mayores, concordando con el perfil de mortalidad de Latinoamérica del final de siglo XX (gráficos 1 y 2).

Al calcular el incremento o reducción del índice AEVP según la categoría de enfermedades del sistema circulatorio, se observó para los hombres

de Bogotá y Cali la mayor reducción del indicador por este grupo de causas, mientras que los hombres de Medellín difieren con las causas latinoamericanas de final de siglo, al

evidenciar una mayor reducción de AEVP en la categoría de todas las demás causas. Las mujeres comparten la misma situación que los hombres (gráficos 3 y 4).

Gráfico 7. AEVP en las mujeres desde la edad de 60 hasta 85 años por subcategorías de causas de muerte, según lista 6/67 OPS. Bogotá, Medellín y Cali. 1999-2001 y 2007-2009.



Fuente: DANE. Registros de defunciones de Estadísticas Vitales y Proyecciones de población. Elaboración propia usando la rutina PAS (AEVP-6085).

Si siguiendo el análisis de CELADE y otros, las reducciones más significativas en las tasas estandarizadas a nivel regional se observaron en las subcategorías de enfermedades cerebrovasculares, así como en enfermedades isquémicas coronarias. Para el caso de las tres ciudades colombianas analizadas con los AEVP en el primer decenio del siglo XXI, se confirma la tendencia en la reducción de enfermedades isquémicas

del corazón y cerebrovasculares a pesar de que continúan siendo las dos principales causas de mortalidad, lo que indica el esfuerzo por parte del sector salud en la reducción de los mismos en los últimos años o el mejoramiento en la calidad de los datos.

Para Latinoamérica, la tasa estandarizada de mortalidad debida a todas las demás causas en adul-

tos mayores aumentó entre 1980 y 2000. Dentro de este grupo adquiere relevancia la diabetes mellitus, que se incrementó notablemente en el periodo descrito (CELADE et al., 2003). Para Bogotá, Cali y Medellín, se presenta como la segunda categoría que más años de esperanza de vida aportó debido probablemente a los controles hechos por el sector salud. Aunque la diabetes mellitus se reporta como la cuarta subcategoría con mayor índice de AEVP, revela que se está afrontando en Colombia el reto observado en Latinoamérica de disminuir la mortalidad por esta causa en la región, pero que este esfuerzo requiere de mayor atención (gráficos 6 y 7).

La tasa estandarizada de mortalidad en América Latina y el Caribe debida a enfermedades neoplásicas aumentó levemente en hombres (4%) y descendió también levemente en mujeres (5%) (CELADE, CEPAL et al., 2003). En las tres ciudades estudiadas, las neoplasias ocupan el tercer grupo de grandes causas de muerte en importancia, observándose un incremento mayor en hombres que en mujeres durante el decenio estudiado según los AEVP (gráficos 1 y 2).

En la población masculina adulta mayor latinoamericana, la tendencia más notoria había sido el aumento de 52% en el riesgo de muerte por cáncer de próstata y de 6% en el de pulmón, incrementos no compensados por la reducción de 25% en el riesgo de muerte por cáncer de estómago y de 1% en las demás neoplasias (CELADE, CEPAL et al., 2003). En los hombres de Bogotá, Cali y Medellín resalta el tumor maligno del estómago como la principal causa del grupo de las neoplasias y con mayor reducción en la pérdida de años de vida, seguido de otros tumores malignos, lo que indica una diferencia con respecto a lo que sucedía en Latinoamérica a finales de

siglo, donde el cáncer de próstata ocupa el segundo o tercer lugar de causas para los hombres (según la ciudad) junto con otros tumores malignos (gráfico 6).

Por su parte, en la población femenina latinoamericana de esta edad, el riesgo de muerte por cáncer de pulmón se incrementó un 25% y el de mama 15%, lo que

fue parcialmente compensado por la reducción del riesgo de muerte por cáncer de estómago de 34%, de útero de 14% y de las demás neoplasias de 3%. En su conjunto, la mortalidad por cáncer en este grupo de edad representaba la segunda causa de mayor magnitud, después de las enfermedades cardiovasculares (CELADE, CEPAL et al., 2003). Para las tres ciudades

de Colombia, las mujeres reportan otros tumores malignos como los causantes de mayor AEVP por neoplasias; sin embargo se observa una distribución relativa en los AEVP sobre varias causas, siendo la muerte por cáncer de pulmón una de las causas representativas con mayor incremento en Medellín y Cali sobre Bogotá (gráfico 7).

Conclusiones

Finalmente, este comparativo entre Bogotá, Cali y Medellín permitió reflexionar sobre la mortalidad de los adultos mayores de estas tres ciudades durante 1999 y 2009 por sexo; desagregando las causas de mortalidad de acuerdo con la propuesta de agrupación de la lista 6/67 de la OPS y midiendo el nivel de contribución de las mismas a los años de esperanza de vida perdidos de los adultos mayores.

Se pudo observar que las enfermedades isquémicas del corazón y las enfermedades cerebrovasculares son las principales causas de fallecimiento en la población adulta mayor de las tres ciudades, ubicándolas en la tercera fase de la transición epidemiológica; aunque se

evidenció una tendencia a la reducción por estas causas, indicando un esfuerzo por parte del sector salud colombiano en la reducción de los mismos en el periodo de referencia, al igual que la diabetes mellitus, la cual cobraba importancia en Latinoamérica a finales del siglo XX.

Se observó un incremento en el nivel de la mortalidad por efecto de las enfermedades infecciosas en las tres ciudades comparadas, específicamente por infecciones respiratorias agudas, junto con los síntomas, signos y afecciones mal definidas, generando un llamado de atención en la prevención por oleadas invernales y a las distintas fuentes notificadoras de hechos vitales en la debida capacitación del

personal notificador, mejorando el nivel de omisión o no reporte de la información de las estadísticas vitales de defunciones en cada ciudad.

Se recomienda a los distintos organismos del Estado mejorar las políticas de acceso a exámenes diagnósticos preventivos, por ejemplo para la detección del cáncer de mama de la mujer y tumores de estómago; también la mejora en la atención primaria como en el caso de la diabetes mellitus y las infecciones respiratorias e hipertensivas, incluyendo un mayor acceso a los medicamentos. También direccionar los programas en la reducción de la incidencia de la mortalidad, con apoyo interinstitucional según las necesidades propias de cada ciudad.

Bibliografía

Arriaga, Eduardo. (1989). Changing trends in mortality decline during the last decades. *Differential Mortality*, editado por Ruzicka et al., Oxford Press, pp. 105-130.

— (1989). Measuring and explaining the change of temporary life ex-

pectancies. *Demography*, 21 (1), February.

— (1997). El uso de años de vida perdidos para medir el nivel y cambio de la mortalidad. *Notas de Población*, año XXIV, núm. 63, Latin American Demographic Center, Santiago.

— (2011). *Análisis demográfico de la mortalidad*. Inédito.

Cardona, Doris y Peláez, Enrique (2010). Mortalidad y morbilidad atendida de personas mayores Medellín, Colombia. Trabajo presentado en el IV Congreso de la Asocia-

ción Latinoamericana de Población, realizado en La Habana, Cuba, del 16 al 19 de noviembre de 2010.

CELADE, CEPAL, UNFPA, OPS, OIT, BID, BM y NU (2003). *Las personas mayores en América Latina y el Caribe: diagnóstico sobre la situación y las políticas*. Santiago de Chile: CELADE/CEPAL.

CELADE y CEPAL (2006). *La protección social de cara al futuro: acceso, financiamiento y solidaridad*. Santiago de Chile: CELADE/CEPAL.

COLOMBIA y DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA (2012). *Proyecciones de población: proyecciones municipales 2005-2011*. Bogotá D. C.: DANE. Disponible en <http://dane.gov.co>.

— (2011). *Registros de defunción, Medellín 1999-2009*. Bogotá D. C.: DANE.

DANE y UNFPA (2011). *Evaluación de la Cabalidad de las Estadísticas Vitales. Informe final*. 2011. Inédito.

Del Popolo, Fabiana (2001). *Características sociodemográficas y socioeconómicas de las personas de edad en América Latina*. Serie Población y Desarrollo, núm.19. Santiago de Chile, CEPAL/CELADE.

Fundación Santillana para Iberoamérica et al. (2001). *Agenda sobre Envejecimiento Colombia siglo XXI – Memorias de un Foro y Bitácora para Futuros Debates*. Octubre 1 de 2001.

González, C. A. y Ham-Chande R. (2007). *Funcionalidad y salud: una tipología del envejecimiento en México*. *Salud Pública Mex.*, 49 supl. (4): S448-S458.

Naciones Unidas (2002). *Informe de la Segunda Asamblea Mundial sobre el Envejecimiento*. Madrid: NU.

Omran, A. (1971). The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *Millbank Mem Fund. Q.*, 49(4): 509-538.

— (1998). The Epidemiologic Transition theory revisited thirty years later. *World Health Statistics Quarterly*, 1998; 51(2, 3, 4): 99-119.

Organización Panamericana de la Salud (1999). LISTA OPS 6/67 (CIE/10, 7 de julio de 1999). Dr. Roberto Becker (WDC) O.P.S. programa especial de análisis de salud.

Organización Mundial de la Salud (2008). *Informe sobre la salud en el mundo 2008: La atención primaria de salud, más necesaria que nunca*. Ginebra.

Salgado de Snyder, N. y Wong, R. (2007). Género y Pobreza: Determinantes de la salud en la vejez. *Salud Pública de México*, 4, núm. Su4: 515-521. Instituto Nacional de Salud Pública, Cuernavaca, México.

Vallin, J. (1994). *La Demografía*. Santiago de Chile, LC/DEM/G.147, Serie E, núm. 41 CEPAL/CELADE.

Villa, M. y Rivadeneira, L. (2000). El proceso de envejecimiento de la población de América Latina y el Caribe: una expresión de la transición demográfica [en línea]. Serie seminarios y conferencias núm. 2, Santiago de Chile: CEPAL/CELADE, Dirección electrónica:

http://www.eclac.cl/publicaciones/xml/4/5604/lcl1399e_IndS1.pdf
Consulta: 11 de agosto de 2012.

Villanueva, L. (2000). Sobre el envejecimiento: una perspectiva integral. *Rev. Hosp. Gral. Dr. M Gea González*, 3(3): 107-114.

Zúñiga, E. y Vega, D. (2004). *Envejecimiento de la población de México. Reto del siglo XXI*. México DF: CONAPO.

A vibrant scene of children playing in a public fountain. The children are of various ages and are splashing water. They are wearing colorful clothing like striped shirts, red dresses, and blue tank tops. The fountain is set in a park-like area with trees and a paved ground. The overall atmosphere is joyful and active.

Análisis espacial multitemporal de la tasa de mortalidad infantil municipal en Colombia

Javier Sebastián Ruiz Santacruz
Carlos Alberto Durán Gil

Análisis espacial multitemporal de la tasa de mortalidad infantil municipal en Colombia

Javier Sebastián Ruiz Santacruz: Estadístico de la Universidad Nacional de Colombia, magíster en Demografía de la Universidad de Estocolmo en Suecia. Asesor de la Dirección de Censos y Demografía. Correo electrónico: jsruiz@dane.gov.co

Carlos Alberto Durán Gil: Ingeniero Catastral y Geodesta de la Universidad Distrital Francisco José de Caldas. Profesional Universitario del Grupo de Investigación y Desarrollo de la Dirección de Geoestadística del DANE. Correo electrónico: cadurang@dane.gov.co

Fecha de recepción: 13 de junio de 2013
Fecha de aceptación: 9 de octubre de 2013

Resumen: este artículo describe el desarrollo del análisis espacial de la Tasa de Mortalidad Infantil (TMI) municipal en Colombia, obtenida a través de las Estadísticas Vitales (EEVV) por el Departamento Administrativo Nacional de Estadística (DANE), para distintos periodos (2004-2007-2010) con el fin de realizar un análisis que contemple el espacio como factor explicativo del comportamiento de la TMI.

Las etapas del proceso comienzan por el Análisis Exploratorio de Datos Espaciales (AEDE), el cual estará enfocado a examinar la distribución y heterogeneidad de los datos de manera preliminar a través de estadística, soportada con representación cartográfica temática de la TMI a nivel municipal. A continuación se realizará el Análisis de Patrones y Autocorrelación Espacial, para determinar agrupaciones o concentraciones de valores altos o bajos y su probabilidad de propagación, con base en estadística espacial. Finalmente se desarrollará el Estudio Geoestadístico para realizar estimaciones de la tasa a través de una superficie de tendencia continua.

La estadística espacial muestra que la TMI presenta concentración de valores y su ubicación en determinadas zonas del país, corroborada por la autocorrelación espacial. La estimación de la tasa a través de geoestadística muestra diferencias entre los periodos. Los errores de la interpolación disminuyen con el pasar de los años, debido en cierta medida a la mejora en el registro de datos, lo que garantiza una menor incertidumbre en lugares donde la información no presenta una calidad verificable.

Aunque los resultados no son absolutos, la meta de explicar el comportamiento y evolución de la TMI a través de análisis espacial se cumple en la representación contextualizada del fenómeno, generando un valor agregado fundamental que abre un abanico de posibilidades para el desarrollo de diversos estudios a escalas diferentes.

Palabras clave: tasa de Mortalidad Infantil, Estadísticas Vitales, Análisis Espacial, Estadística Espacial, Geoestadística.

Abstract: *this article describes the development of spatial analysis of the Infant Mortality Rate (IMR) Colombia municipal obtained through Vital Statistics (VS) by the National Administrative Department of Statistics (DANE), for different periods (2004-2007-2010) in order to perform an analysis that considers the space as a factor explaining the behavior of the TMI.*

The stages begin with the Exploratory Spatial Data Analysis (ESDA), which will be focused to examine the distribution and heterogeneity of a preliminary data through statistics, supported with thematic mapping of IMR at municipal level. Then be held on Pattern Analysis and Spatial Autocorrelation to determine clusters or concentrations of high or low values and their likelihood of propagation, based on spatial statistics. Finally develop Geostatistical Study for rate estimates through a continuous trend surface.

Spatial statistics has shown that the IMR concentration values and their location in certain parts of the country, supported by spatial autocorrelation. The estimated rate through geostatistics shows differences between periods. Interpolation errors decrease with the passing of the years, due in part to improved data recording, ensuring less uncertainty in places where the information is submitted verifiable quality.

Although the results are not absolute, the goal of explaining the behavior and evolution of IMR through spatial analysis is fulfilled in the contextual representation of the phenomenon, generating added value key that opens a range of possibilities for the development of various studies at different scales.

Keywords: *infant Mortality Rate, Vital Statistics, Spatial Analysis, Spatial Statistics, Geostatistics.*

1. Introducción

La mortalidad infantil es un indicador utilizado para monitorear el estado de salud de una población, y principalmente para evaluar las condiciones de vida de la población, ya que se considera que refleja las condiciones sanitarias, las prácticas dietéticas y el estado nutricional de la población (Céspedes, 2008). El DANE, a través de la Dirección de Censos y Demografía (DCD), es el responsable de estimar la TMI anualmente para cumplir con los Objetivos de Desarrollo del Milenio, siendo parte de la meta 4A de las Naciones Unidas. En Colombia, la TMI es requerida por varias instituciones estatales para hacer seguimiento a los programas de gobierno que se implementan desde las secretarías de planeación, tanto departamentales como municipales, y desde los distintos ministerios e institutos.

La TMI ha entrado en un proceso de revisión permanente y los resultados se han derivado de la construcción de un modelo estadístico construido por el DANE, que estima el cambio anual. Las descripciones encontradas acerca de la TMI están enfocadas con base en niveles estimados para todos los años; es obtenida mediante el modelo de estadístico de estimación que da cuenta del cambio de la misma de un año a otro. Los análisis realizados hasta el momento por la DCD se refieren a modelamientos estadísticos para su cálculo y tendencias, pero no se tiene una concepción espacial del fenómeno que permita describir patrones o comportamientos regionales. Los métodos de análisis espacial empleados permiten dar un enfoque amplio al estudio, incorporando el espacio como una dimensión explicativa, tanto en la visualización concreta del fenómeno, y fundamentalmente, como factor influyente en la tendencia y comportamiento de la misma tasa. De acuerdo con la Primera Ley de

la Geografía¹, la similitud locacional y de valores influye en comportamientos asociativos de las tasas de mortalidad, como se observó en otros estudios relacionados en Estados Unidos (Kim, 2011), México (Álvarez, 2009) y Argentina (Ponce, 2012). Lo anterior conlleva a la necesidad de realizar un análisis que permita determinar a escala general la evolución, las tendencias, los patrones espaciales y las estimaciones de la tasa, generando resultados con valor agregado para la evaluación integral del fenómeno sociodemográfico.

Las decisiones adoptadas para la estimación de la tasa y para el mejoramiento de la calidad de la misma requieren el análisis del comportamiento para cada uno de los años de la TMI, por esto se hace indispensable el inicio de un análisis detallado de los determinantes de la mortalidad infantil (MI), empezando por su correlación espacial dentro del territorio, y así establecer herramientas suficientes para esclarecer el

¹ Principio enunciado por Waldo Tobler, en el cual afirma que las cosas que están más cercanas entre sí poseen una similitud mayor que aquellas ubicadas a una distancia mayor.

tratamiento que debe hacerse para obtener mejores estimaciones. El plan de mejoramiento de las Estadísticas Vitales será articulado con los resultados de este análisis espacial para la toma de decisiones, dentro del Proyecto de Fortalecimiento del Sistema de Registro Civil y Estadísticas Vitales de Colombia 2012-2014, siendo un insumo conceptual en esta área.

Este documento pretende, entonces, realizar un análisis espacial multitemporal de la TMI, con base en los valores generados por las defunciones y nacimientos, obtenidas de las estadísticas vitales de Colombia. La información seleccionada para el desarrollo del artículo son los años 2004, 2007 y 2010, con el fin de realizar un estudio longitudinal que permita visualizar de manera más idónea los cambios en el tiempo.

2. Descripción de la información

La información analizada es aquella calculada directamente, usando los nacimientos y las defunciones de las Estadísticas Vitales (EEVV). En algunos años no se presentan nacimientos y defunciones en algunos municipios (lo que es totalmente factible) y en ellos se generan valores no disponibles en el cálculo de las tasas. Además, como la información está sin ajuste, existen datos atípicos que superan los niveles considerados como aceptables de la tasa y esto, a su vez, tiene repercusiones en los estadísticos básicos como la varianza, la media y la mediana. Es necesario aclarar que la TMI calculada en términos demográficos como una tasa edad-periodo se denomina Índice Sintético de Mortalidad Infantil, ya que los cálculos se hacen con los datos de las defunciones de un año sobre los nacimientos del mismo. No es posible calcular una tasa más refi-

nada debido al rezago de tiempo de entrega que tienen las EEVV y su importancia en la actualización en términos de la política pública.

La fórmula de cálculo para la tasa se define de la siguiente manera:

$$TMI = \frac{dx}{Nx}$$

Donde dx son las defunciones registradas en el sistema de EEVV en el año x y Nx son los nacimientos registrados en el sistema de EEVV en el mismo año. Esta tasa se entrega por cada 1000 nacidos vivos.

La escogencia de los años de estadísticas vitales para el análisis se debió a procesos que estaban relacionados con los cambios en su estructura. Según el documento «Las estadísticas vitales en Colombia» (DANE, 2011), en el año de 1998, el Departamento Administrativo Nacional de Estadística (DANE) lleva el registro de los nacidos vivos y las defunciones del país, ya que anteriormente era la Registraduría Nacional del Estado Civil quien llevaba este proceso. No hubo mayor intervención en el manejo de la captación hasta que en el año 2002 se crea la Comisión Intersectorial de Estadísticas Vitales, conformada principalmente por el Ministerio de Salud, el Instituto Nacional de Salud (INS), la Registraduría Nacional del Estado Civil, la Superintendencia de Notariado y Registro, el Instituto de Medicina Legal y Ciencias Forenses, el Instituto Colombiano de Bienestar Familiar (ICBF), el Departamento Nacional de Planeación (DNP), Profamilia y el Cuerpo Técnico de Investigación de la Fiscalía (CTI). De esta comisión formada en el segundo semestre del referido año, no era de esperarse que presentara resultados en el año siguiente. Por esto, se tomaron las estadísticas vitales de 2004 y así poder observar su posible cambio. En el año 2005 se realiza el Censo General, que indica un cambio en la línea

base; es decir, estimada a partir del censo, que repercute en las estimaciones posteriores, al registrar un descenso que obliga una reestimación de las tasas de los años 2006, 2007, 2008 y 2009; sin embargo, hasta ese momento no hay una intervención en la estructura del registro. Solo hasta 2007 se cambia el formulario y la estructura del mismo con el fin de mejorar el fraseo de las preguntas y su organización. Ya en el año 2008 entra a funcionar el sistema de Registro Único de Afiliación (RUAF), el cual permite la captura de los hechos vitales vía web, diseñado para poder generar un incremento en la cobertura, ya que parte de la TMI también se debe a un mal registro de los nacimientos y las defunciones.

En general, los problemas presentados en las EEVV son de cobertura (registros no reportados, registros perdidos, registros tardíos o extemporáneos y no registro) y de calidad (diligenciamiento completo del certificado o integralidad, coherencia entre las variables del certificado o validez, y calidad de certificación de la causa de muerte). El informe final de la Evaluación de la Cabalidad de las Estadísticas Vitales (DANE, 2011), cuyo objetivo fue identificar, ubicar y caracterizar los principales problemas hacia una mejora del registro vital, reconoció una no cobertura total nacional de las defunciones del 18,2%, y para los nacimientos una no cobertura del 13,6%, ambas ajustadas por no reporte. A nivel departamental, la no cobertura tiene una variación considerable, teniendo, por ejemplo, departamentos donde la no cobertura es de 64,5% como La Guajira, o 58,3% en Amazonas, hasta departamentos cuya no cobertura es de 6,8%, como es el caso de Boyacá, y 8,8% para Meta, según los datos del informe.

En cuanto a la información geográfica, la información de cada año fue asociada a su respectiva vigencia

del Marco Geoestadístico Nacional a nivel municipal, para garantizar la relación idónea de la tasa al municipio correspondiente, ya que a través del tiempo es posible la presencia y configuración de nuevos municipios, y es fundamental asegurar que la zona de influencia, en este caso el territorio, cuente con la mejor cobertura posible.

3. Metodología, análisis y resultados

La metodología del análisis espacial de la TMI está basada en el concepto PPDAC: Problema, Plan, Datos, Análisis y Conclusiones. Este concepto, desarrollado por Smith & Goodchild, examina los procesos de análisis en un contexto amplio de metodología analítica, a través de etapas iterativas, desarrolladas para este trabajo de la siguiente manera:

- Problema: Explicar la TMI desde el contexto espacial.
- Plan: Análisis Espacial de la Tasa de Mortalidad Infantil.
- Datos: Información de TMI para los años 2004, 2007, 2010.
- Análisis: Análisis Exploratorio de Datos Espaciales, Análisis de Patrones y Autocorrelación Espacial, Estudio Geoestadístico.
- Conclusiones.

3.1. Análisis Exploratorio de Datos Espaciales (AEDE)

El Análisis Exploratorio de Datos Espaciales (AEDE) es el conjunto de técnicas que describen y visualizan las distribuciones espaciales de los fenómenos. Aunque es un estudio exploratorio con carácter descriptivo, permite determinar estructuras espaciales en las va-

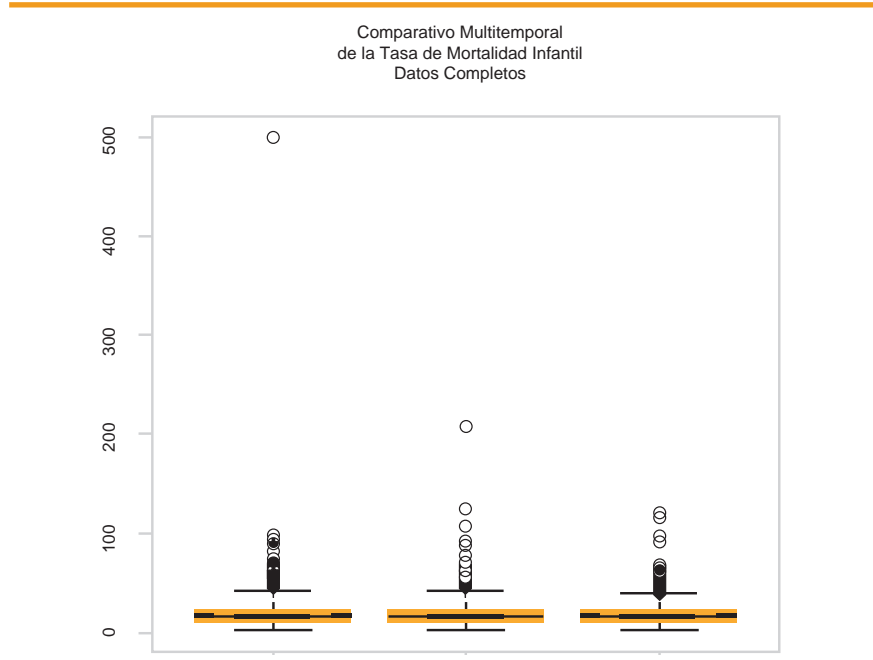
riables, para formular hipótesis previas (Chasco, 2003).

En este proceso se describe y visualiza la distribución espacial de la tasa con base principalmente en los datos. Esto se hace a través de la representación de distribuciones (diagrama de caja) y de la repre-

sentación de heterogeneidad espacial (análisis estadístico descriptivo y mapas temáticos).

Un primer análisis a través del diagrama de caja muestra los datos atípicos, pero no permite observar con claridad el comportamiento de la tendencia de la TMI.

Gráfico 1. Caja de datos completo.



Fuente: elaboración propia.

La regla para la determinación de dichos datos fue seleccionar aquellos puntos estandarizados que se encontraban a más de tres desviaciones estándar y para los cuales se les iba a dar otro tratamiento. Los histogramas generados no se pueden ver en detalle debido a la escala de los datos atípicos; sin embargo se puede observar que la mayoría de las distribuciones son de cola pesada a la derecha; es decir, la distribución de la TMI no presenta un comportamiento de normalidad, que se debe dar como condición subyacente en geoestadística, para minimizar el grado de incertidumbre de las estimaciones.

El tratamiento dado a los datos atípicos fue el de removerlos de la in-

formación primaria; esto con el fin de realizar análisis estadísticos sin este «ruido blanco», y posteriormente estimar estos valores a través de interpolación espacial bajo criterios geoestadísticos. Ahora, esto ocurre si se asume que todos los datos atípicos son generados por cualquiera de las dos fuentes de omisión, bien sea de las defunciones o los nacimientos.

Un primer análisis estadístico exploratorio de los datos sin datos atípicos descritos en la gráfica 3 muestra una tendencia decreciente de la TMI en las EEVV. En el gráfico se observan las muescas hechas por el programa alrededor de la mediana, que son los intervalos de confianza para dicha medida

de tendencia central; esto con el fin de realizar evaluaciones. Ahora se siguen viendo datos atípicos en los nuevos *boxplot*, pero con la aclaración de que todos los datos remanentes del proceso anterior se encuentran dentro de tres desviaciones estándar.

Con la eliminación de los datos atípicos obtenemos histogramas de frecuencia que pueden observarse

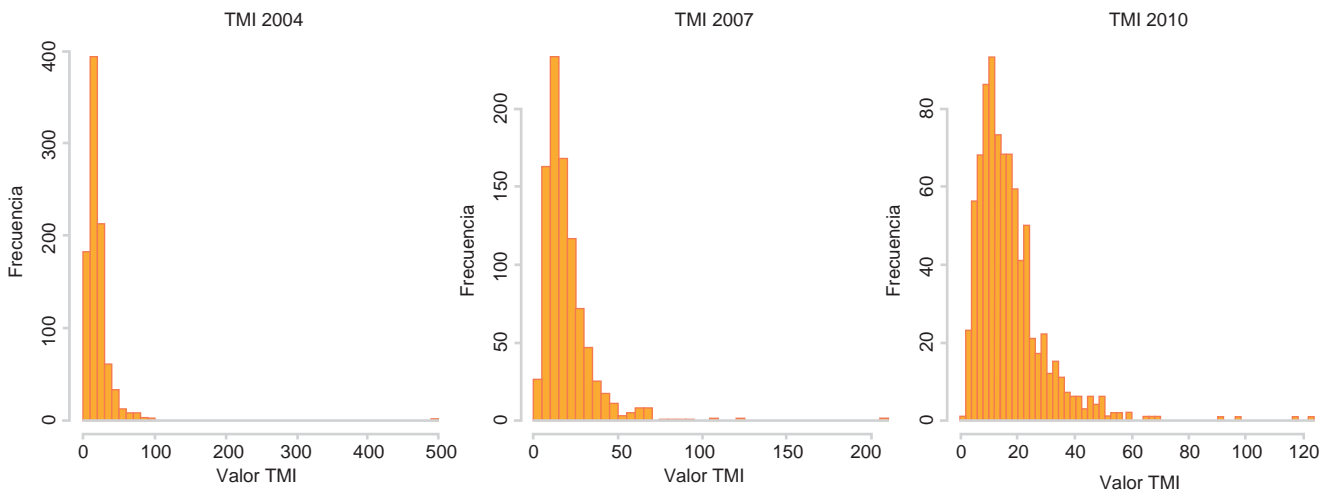
con mejor detalle y confirmar cómo los datos se concentran en los valores bajos de la tasa a medida que pasa el tiempo.

La representación primaria de la TMI se hace a través de mapas temáticos convencionales, en donde también se identifican los datos atípicos.

Los mapas representan la TMI para cada municipio en cada año, pero

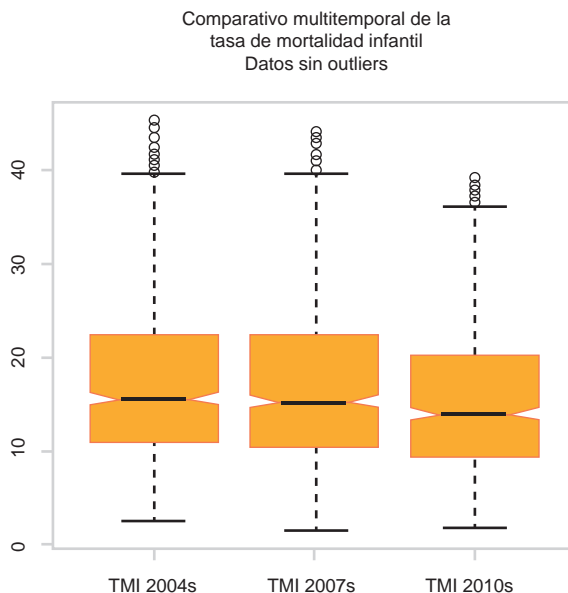
no es lo suficientemente claro visualizar la disminución a través del tiempo. Se marca una tendencia de concentración de valores altos en Vichada, Guainía, Amazonas y Chocó, y otras subregiones como el sur de Córdoba y el Catatumbo. En gran parte de las regiones Andina, y Caribe, además del Piedemonte, prevalecen las tasas medias y bajas.

Gráfico 2. Histogramas datos completos.



Fuente: elaboración propia.

Gráfico 3. Diagrama de caja sin atípicos.

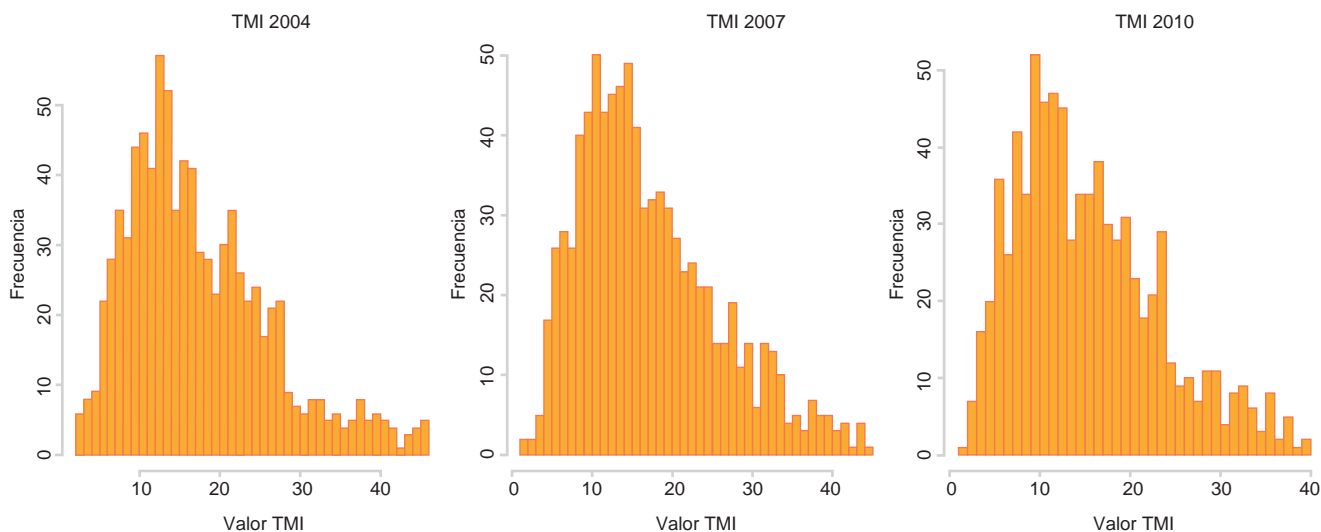


Fuente: elaboración propia.

3.2. Análisis de patrones y autocorrelación espacial

Con base en la información de la TMI ajustada en el paso anterior, se busca identificar posibles patrones o concentraciones de la tasa y su ubicación en el territorio nacional. Para ello se emplean estadísticos espaciales globales y locales que determinan la correspondencia entre los valores de la tasa y su ubicación geográfica. La relación espacial se conceptualizó definiendo una banda de distancia que define la influencia del dato y su entorno, debido al amplio rango de los datos de la tasa y particularmente por la compleja división territorial de los municipios del país, enmarcada en grandes diferencias de áreas y la posición del Archipiélago de San Andrés y Providencia.

Gráfico 4. Histogramas sin atípicos.



Fuente: elaboración propia.

Como primer paso se analizó la autocorrelación espacial, que busca explicar si la cercanía geográfica está relacionada con la similitud en los valores de la tasa. Dicho de otra manera, se busca establecer si la TMI presenta un patrón aleatorio, disperso o agrupado, a través del estadístico global Moran's I^2 , buscando rechazar la hipótesis nula, la cual argumenta que no hay asociación espacial en los valores municipales de la tasa. Si el Índice es cercano a +1.0 indica asociación; en cambio, si es cercano a -1.0 indica dispersión. Los resultados muestran que la TMI presenta patrón de asociación; es decir, autocorrelación espacial, y esta es mayor con el pasar de los años.

Ahora se determina qué tipo de asociación se presenta, ya sea de valores bajos o altos de la tasa, a través del estadístico Getis-Ord General G^3 . Si el valor del Z score es grande y el p-valor es muy pequeño, se rechaza la hipótesis de nula de aleatoriedad (Lentz, 2009). Ahora bien, bajo este precepto, si

el Z score es negativo indica asociación de valores bajos y si es po-

sitivo demuestra lo contrario.

Tabla 1. Estadístico Espacial Moran's I.

Año	Z-score	P-valor	Índice de Moran	Patrón
2004	1,746189	0,080778	0,003575	Agrupado
2007	5,102813	0,000000	0,006921	Agrupado
2010	6,005588	0,000000	0,009748	Agrupado

Fuente: elaboración propia.

Tabla 2. Estadístico espacial Getis-Ord general G.

Año	Z score	P-valor	Asociación
2004	-3,042423	0,002347	Valores bajos
2007	-4,209287	0,000026	Valores bajos
2010	-3,953597	0,000077	Valores bajos

Fuente: elaboración propia.

Los tres años arrojaron concentraciones de valores bajos; sin embargo, esto no excluye que se presenten aglomeraciones de valores altos, ya que una de las desventajas del estadístico es que al agruparse ambos tipos de valores tienden a cancelarse entre sí, dando prevalencia a uno de ellos.

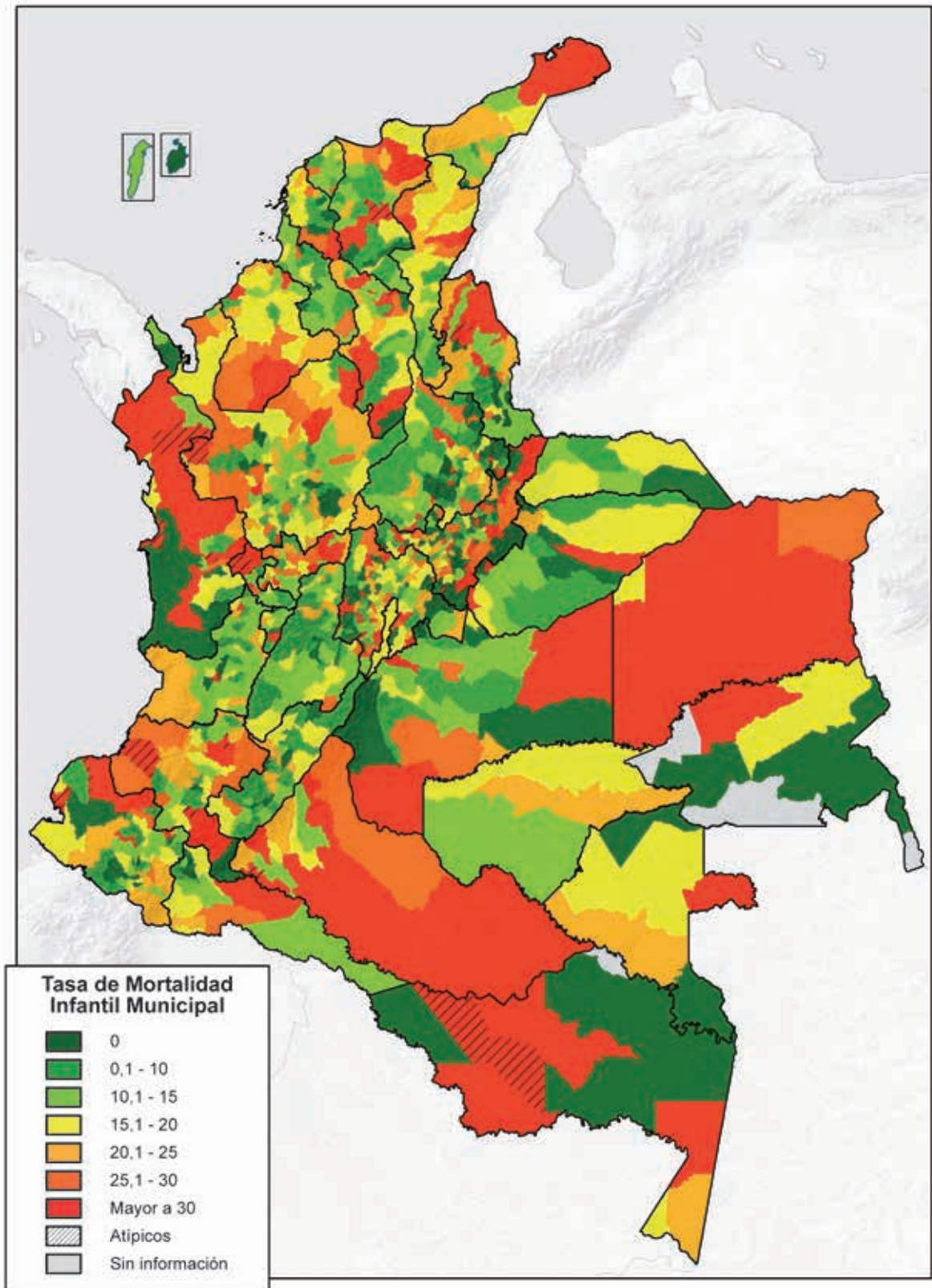
La ubicación espacial de estas asociaciones se hace a través de los denominados puntos calientes (valores altos rodeados de otros valores altos) y fríos (valores bajos rodeados de otros valores bajos), denominado comúnmente análisis Hot-Cold Spot⁴.

² Estadístico espacial global que mide la autocorrelación espacial basado en la ubicación de los elementos y el valor del atributo al mismo tiempo. La magnitud del estadístico depende de la diferencia en los valores del atributo y el número de vecinos con valores similares.

³ Estadístico espacial global que mide el grado de concentración, ya sea, para valores altos o para valores bajos. Al ser un estadístico inferencial se interpreta dentro del contexto de una hipótesis nula, la cual implica que no hay agrupación espacial de los valores.

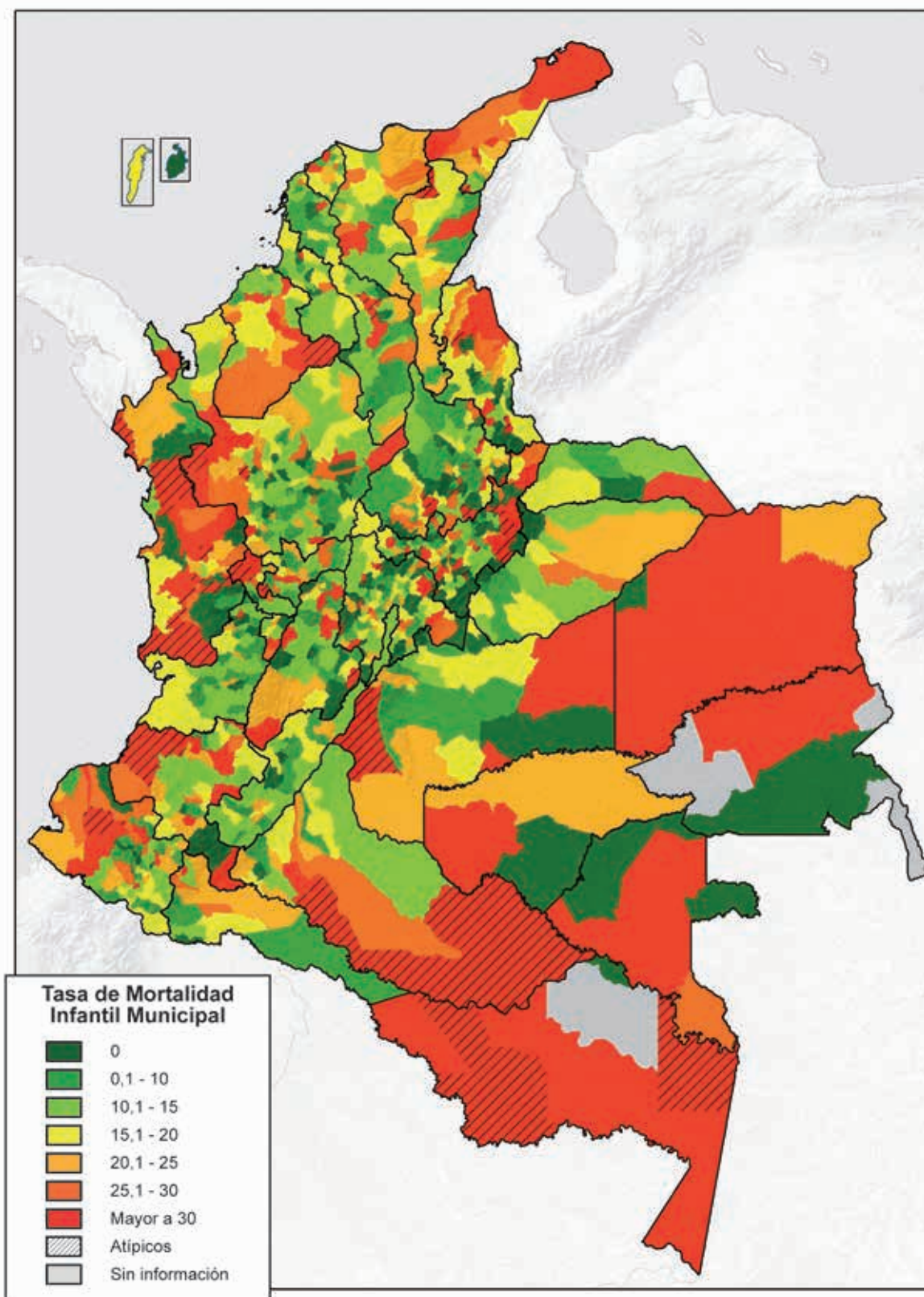
⁴ Este análisis parte de la significancia estadística que debe tener un elemento que tendrá un valor alto que está rodeado por otros elementos altos o viceversa.

Mapa 1. Mapas Temáticos Convencionales (TMI) - 2004.



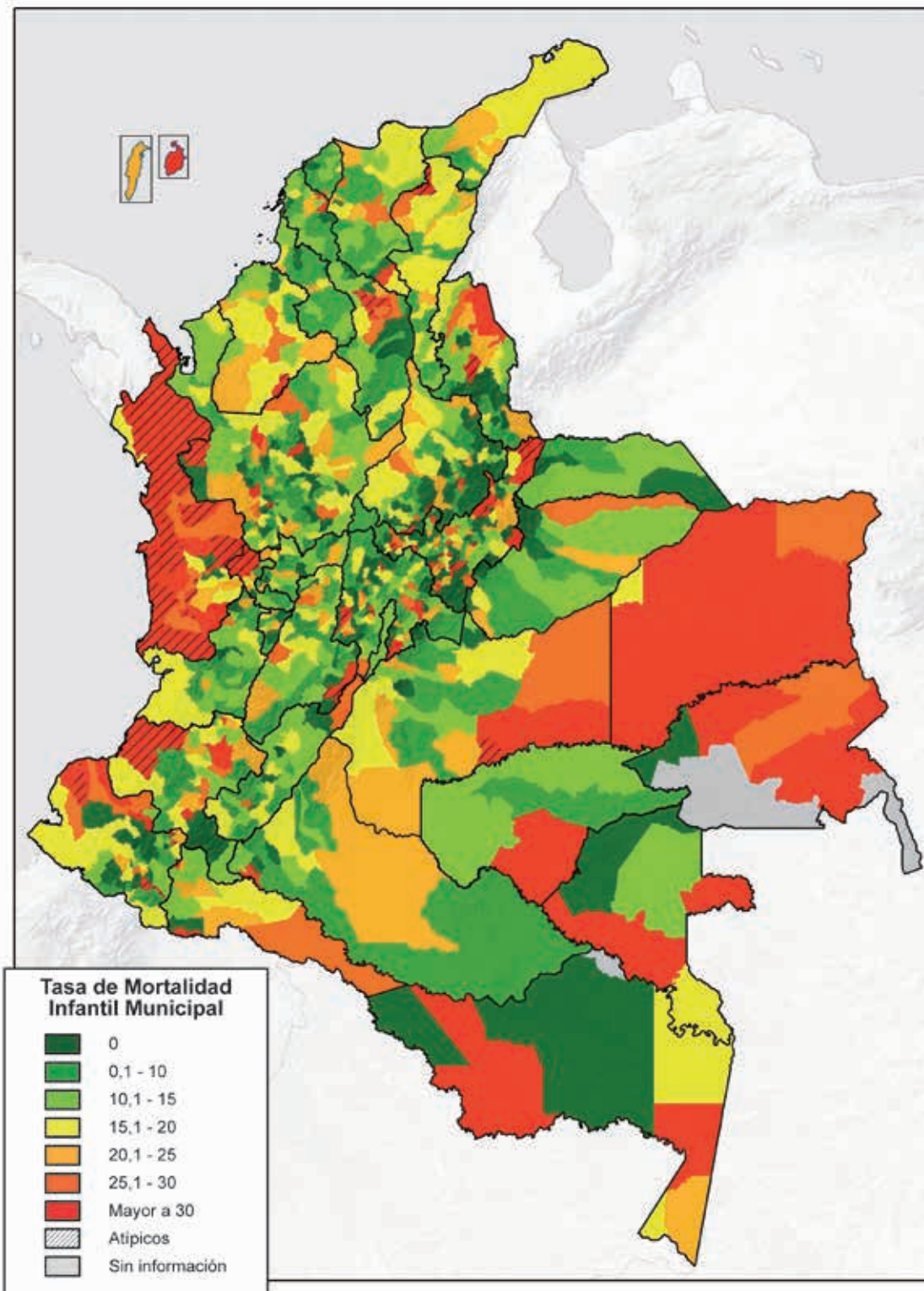
Fuente: elaboración propia.

Mapa 2. Mapas Temáticos Convencionales (TMI) - 2007.



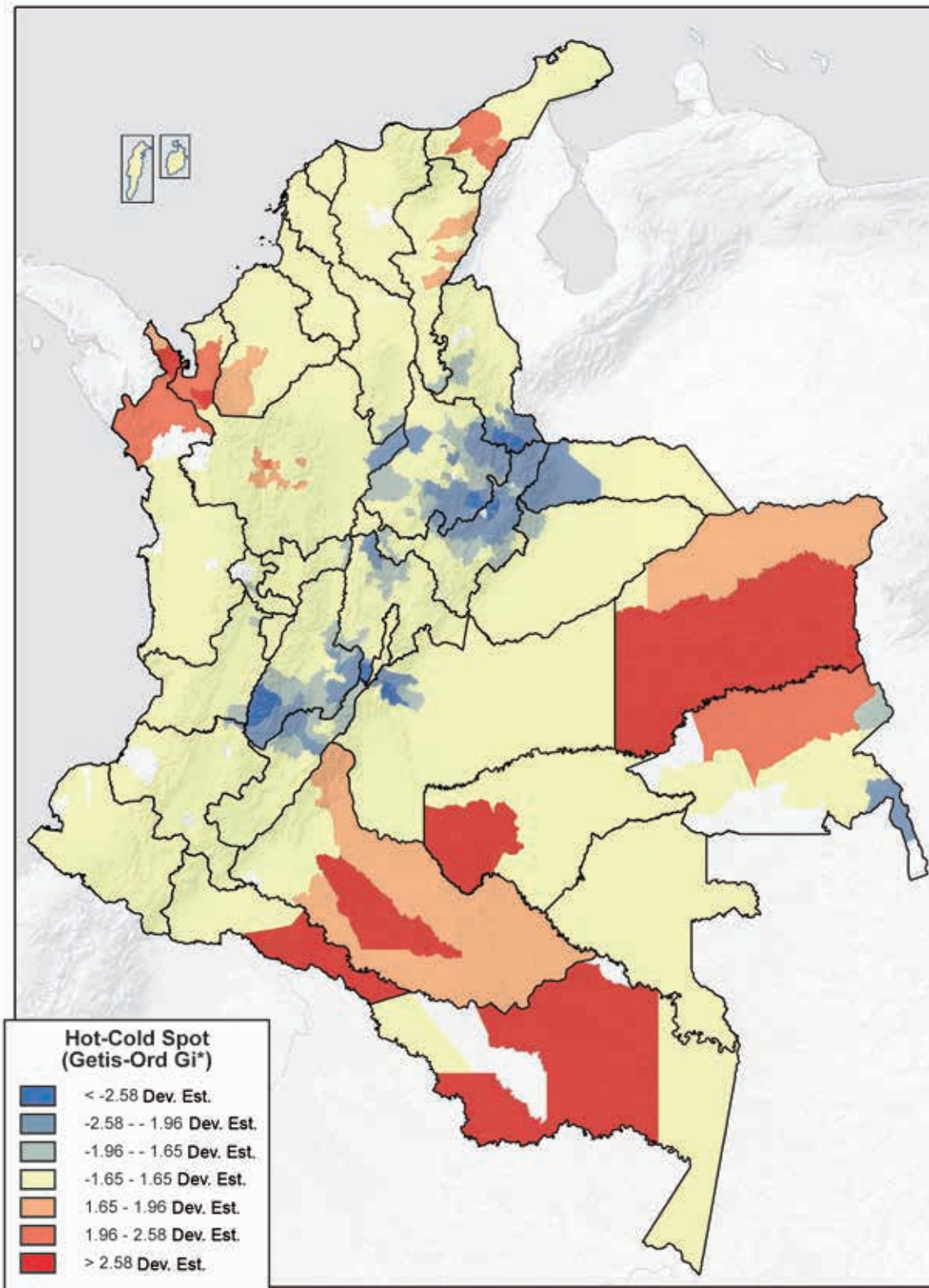
Fuente: elaboración propia.

Mapa 3. Mapas Temáticos Convencionales (TMI) - 2010.



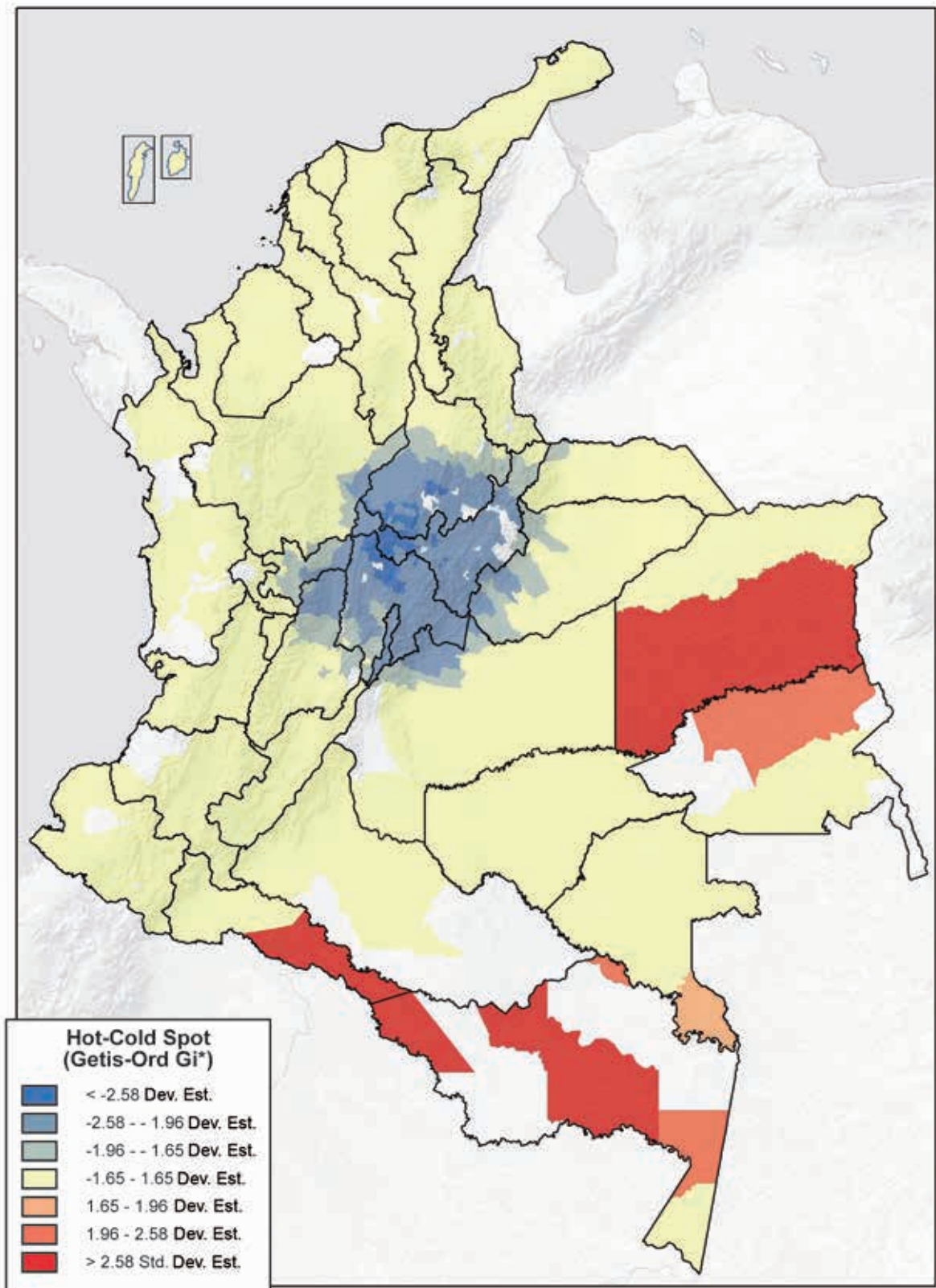
Fuente: elaboración propia.

Mapa 4. Puntos calientes y fríos de la TMI - 2004.



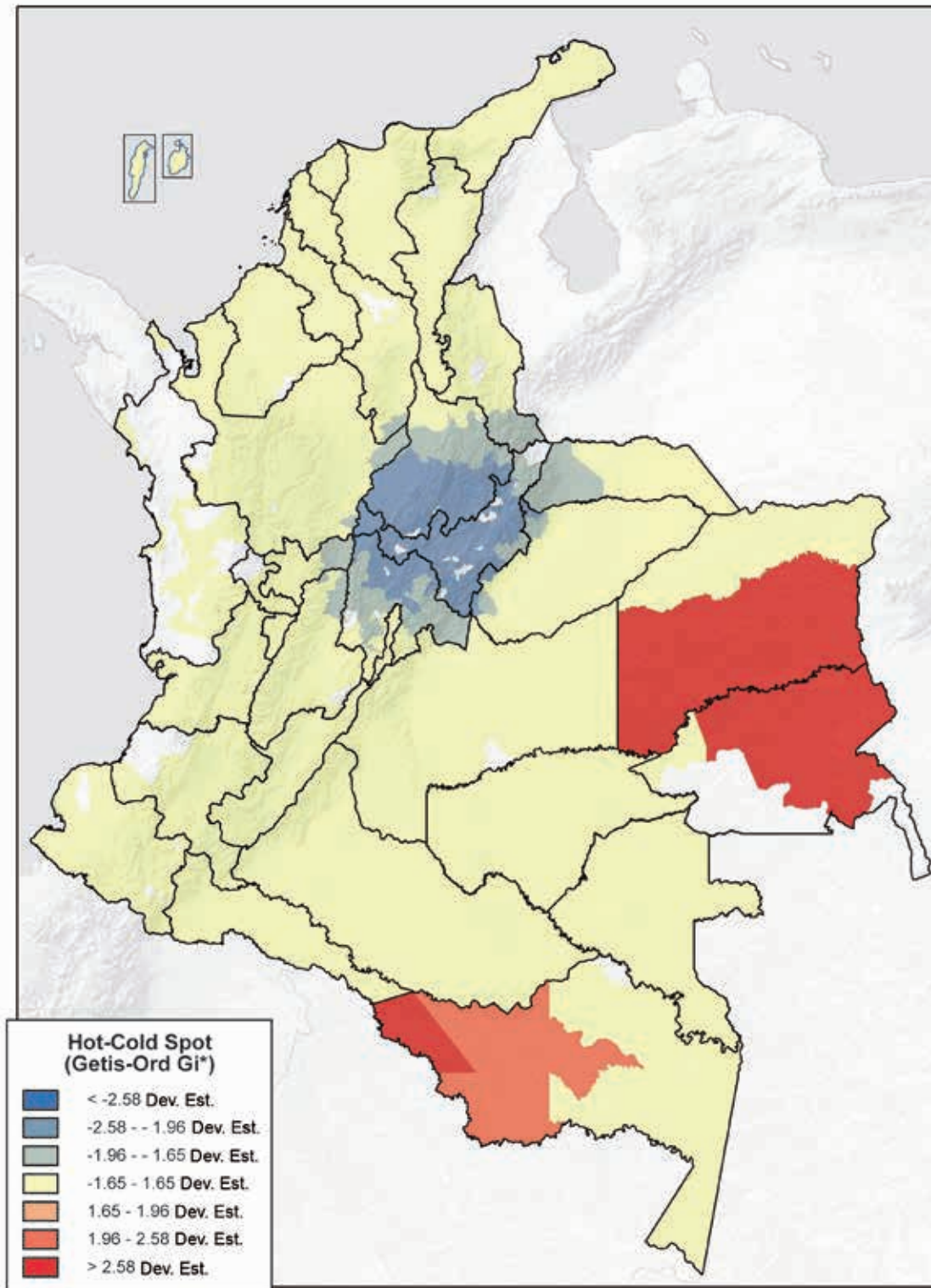
Fuente: elaboración propia.

Mapa 5. Puntos calientes y fríos de la TMI - 2007.



Fuente: elaboración propia.

Mapa 6. Puntos calientes y fríos de la TMI - 2010.



Fuente: elaboración propia.

Los mapas representan una tendencia decreciente de zonas críticas con altos valores de TMI, que se presentan en 2004 principalmente en el Norte de Chocó, Urabá Antioqueño, Sur de La Guajira, Caquetá y Putumayo, y que para 2010 se mantienen particularmente en Vichada y Guainía. Los puntos fríos en 2004 se presentan en ubicaciones particulares y dispersas, entre ellas el Sur del Tolima y el Norte del Huila, Norte de Boyacá y Sur de Santander; en 2007 ya se refleja una concentración clara en Cundinamarca, Boyacá y Santander, permaneciendo en 2010 pero con menor intensidad. Chocó deja de ser relevante en la definición de patrones debido a la presencia de atípicos removidos en el proceso anterior.

Los resultados anteriores muestran diferencias significativas, particularmente entre 2004 y los años posteriores, debido en gran medida a la toma de datos y posiblemente a una reducción en los valores de la tasa, que no es concluyente para determinar posibles contagios de los puntos fríos y calientes a otros vecinos.

3.3. Estudio geoestadístico

La geoestadística permite describir la continuidad espacial de un fenómeno particular, para resaltar la forma en que varía en el territorio

(patrón espacial), además de cuantificar dicha variación en diferentes direcciones (Gallardo, 2006). Esto le permite ser un estimador muy robusto y preciso, según los criterios de probabilidad y cercanía.

Para verificar la normalidad de los datos, se calculan el coeficiente de sesgo y el coeficiente de variación, los cuales explican la distorsión de la distribución de los datos y la representatividad de la media, respectivamente. Para cumplir el

precepto de normalidad y reducir la incertidumbre de las estimaciones, es necesario que el sesgo sea menor a 1 y que la variación sea menor a 200. Si estas condiciones no se cumplen, se recomienda normalizar los datos, ya sea a través de transformaciones de tipo raíz

cuadrada y logarítmica, que para el estudio particular no aplican ya que hay presencia de valores cero. Otra solución es remover otros posibles atípicos (valores extremos) gradualmente, criterio que se aplicó en algunos periodos.

Tabla 3. Coeficiente de Sesgo (CS) y Coeficiente de Variación (CV).

Año	CS original	CV original	CS ajustado	CV ajustado
2004	3,194	84,43	1,036	77,48
2007	26,118	208,78	0,905	72,47
2010	24,102	255,19	0,733	82,99

Fuente: elaboración propia.

Otra condición que deben cumplir los datos para ser estimados a través de geoestadística es la asunción de estacionariedad; es decir, la varianza debe ser igual en las diferentes zonas de estudio. La falta de estacionariedad puede deberse a la existencia de tendencia o gradiente espacial que es mayor al área de estudio (Gallardo, 2006). Si este fenómeno se presenta, se debe remover durante el proceso de estimación, para reducir la media del error, hecho que se dio ya que todos los años presentaron tendencia de segundo orden.

La autocorrelación espacial se ratifica a través del semivariograma experimental, el cual plotea la distancia que separa a los puntos, en este caso los centroides municipales, versus la diferencia al cuadrado de los valores de la tasa.

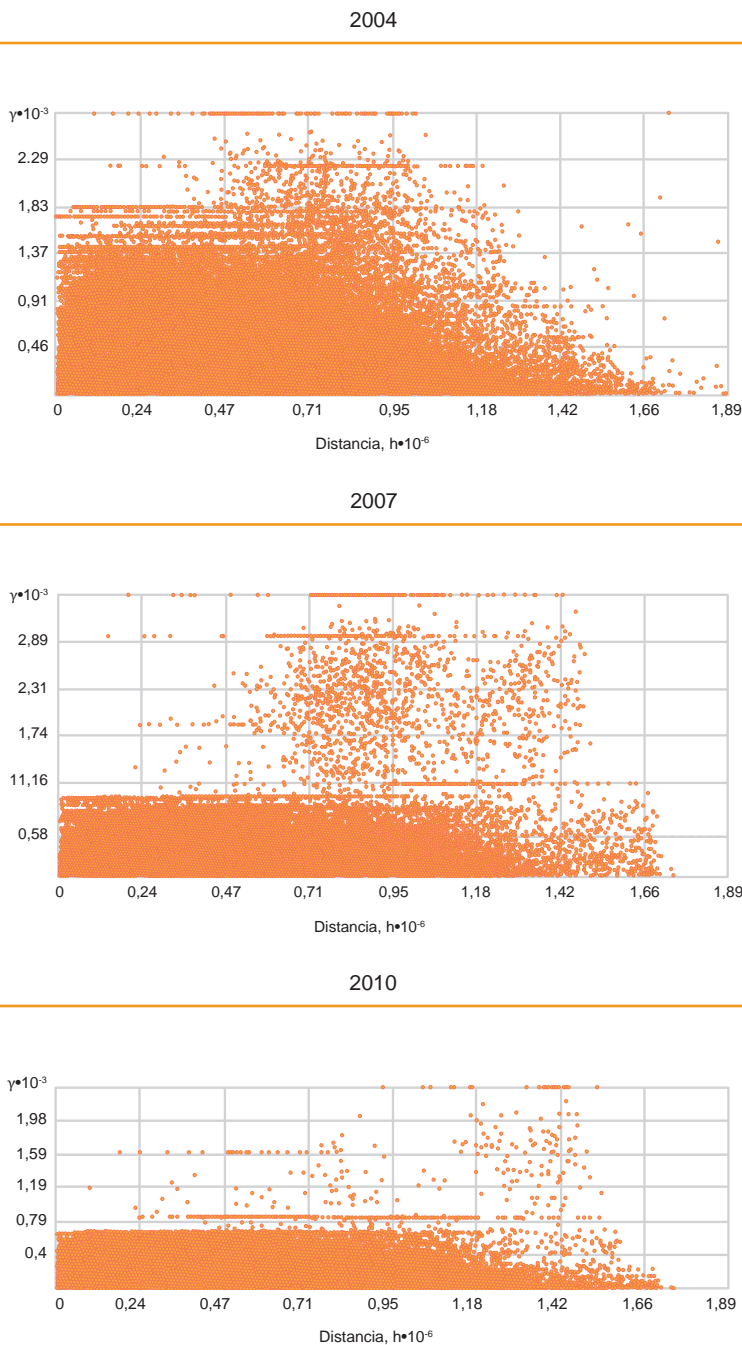
Los tres años presentan autocorrelación espacial, pero en todos ellos se presentan diferencias significativas en la varianza de los datos a distancias muy cortas, además de valores extremos que reflejan ciertas brechas o franjas en la estimación.

Todos los semivariogramas presentan un comportamiento lineal, lo que indica que la TMI es una

variable continua pero su movilidad es considerable, debido a la pendiente que presentan los datos desde el origen hasta la meseta del semivariograma. Esto explica una variación significativa entre los valores a distancias cortas que se observa particularmente en 2004. Con el pasar de los años la variación disminuye indicando una menor volatilidad de la TMI, fenómeno que puede ser explicado en cierta medida por la mejor toma de datos a través de las modificaciones hechas en los últimos periodos.

El proceso para la generación de la superficie de estimación se hará a través de la interpolación por *kriging* ordinario, asumiendo que el valor de la TMI es el mismo para el municipio con base en su centroide, con el fin de minimizar la varianza de los errores. Para ello se parte de analizar los modelos que mejor ajusten los datos, y se elige más adecuado con base en los siguientes requisitos: a) varianza estimada conocida (variabilidad estructural) cercana a la varianza de los datos (*partial sill*); b) menor error medio cuadrático (EMC); c) menor error promedio estándar (EPE); d) error medio cuadrático estandarizado cercano a 1 (EMC_S); e) mayor porcentaje de confiabilidad (calculado como 100 – error máximo).

Gráfico 7. Semivariograma experimental.



Fuente: elaboración propia.

Tabla 4. Parámetros de los modelos escogidos.

Año	Modelo	Varianza	Partial sill	EMC	EPE	EMC_S	Confiabilidad
2004	Esférico	145,54	53,96	12,014	11,745	1,02	92,31
2007	J-Bessel	119,99	102,21	10,721	10,635	1,00	87,61
2010	Pentaesférico	93,51	5,64	9,393	9,446	0,99	92,89

Fuente: elaboración propia.

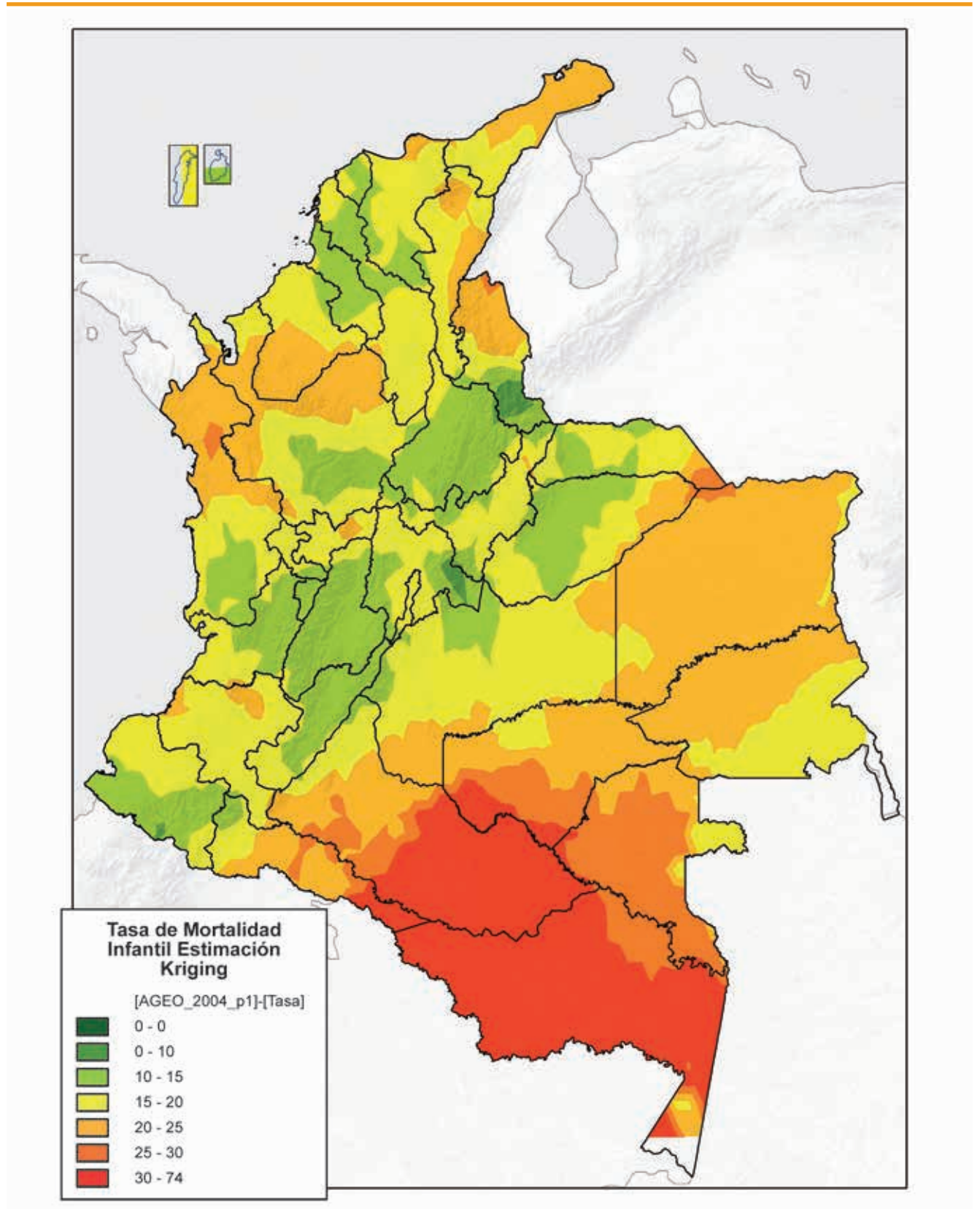
A excepción de la variabilidad estructural, todos los modelos elegidos presentan errores de magnitud similar y confiabilidades altas, lo que representa una adecuada estimación de los valores de la tasa en los tres periodos elegidos. Además, cada año tiene un tipo de modelo específico que ajusta los datos, lo que podría indicar que la condición de los mismos en cada periodo tiene un comportamiento particular. La confiabilidad del modelo en 2007 es menor en comparación a los años anteriores, lo que traduce un mayor error en la predicción, debido a la alta variabilidad de los datos reflejada en la *partial sill*.

A partir de la elección del modelo se generan las superficies de estimación de la TMI para las tres vigencias.

En los mapas de coropletas realizados en el análisis exploratorio, que representan los valores originales de TMI, no se pudo distinguir con claridad tendencias que indicaran aumentos o reducciones en la tasa entre 2004 y 2007. Sin embargo, las superficies de estimación muestran unos cambios relevantes entre periodos, ya que en 2004 prevalecen valores medios, medios-altos y una zona consolidada de valores altos en parte de Caquetá y Amazonas; en 2007 se vislumbran más contrastes, explicados en cierta medida por el mayor grado de incertidumbre del modelo debido a la variabilidad de los datos, presentándose zonas de valores altos en Amazonas, Vichada, Caquetá y La Guajira, y valores bajos particularmente en Boyacá. El año 2010 presenta un patrón más predominante de valores bajos en todo el centro del país y la zona norte, además de gran parte del Amazonas, en discrepancia con Vichada, Guainía, San Andrés y Providencia, y los municipios de la Costa Pacífica.

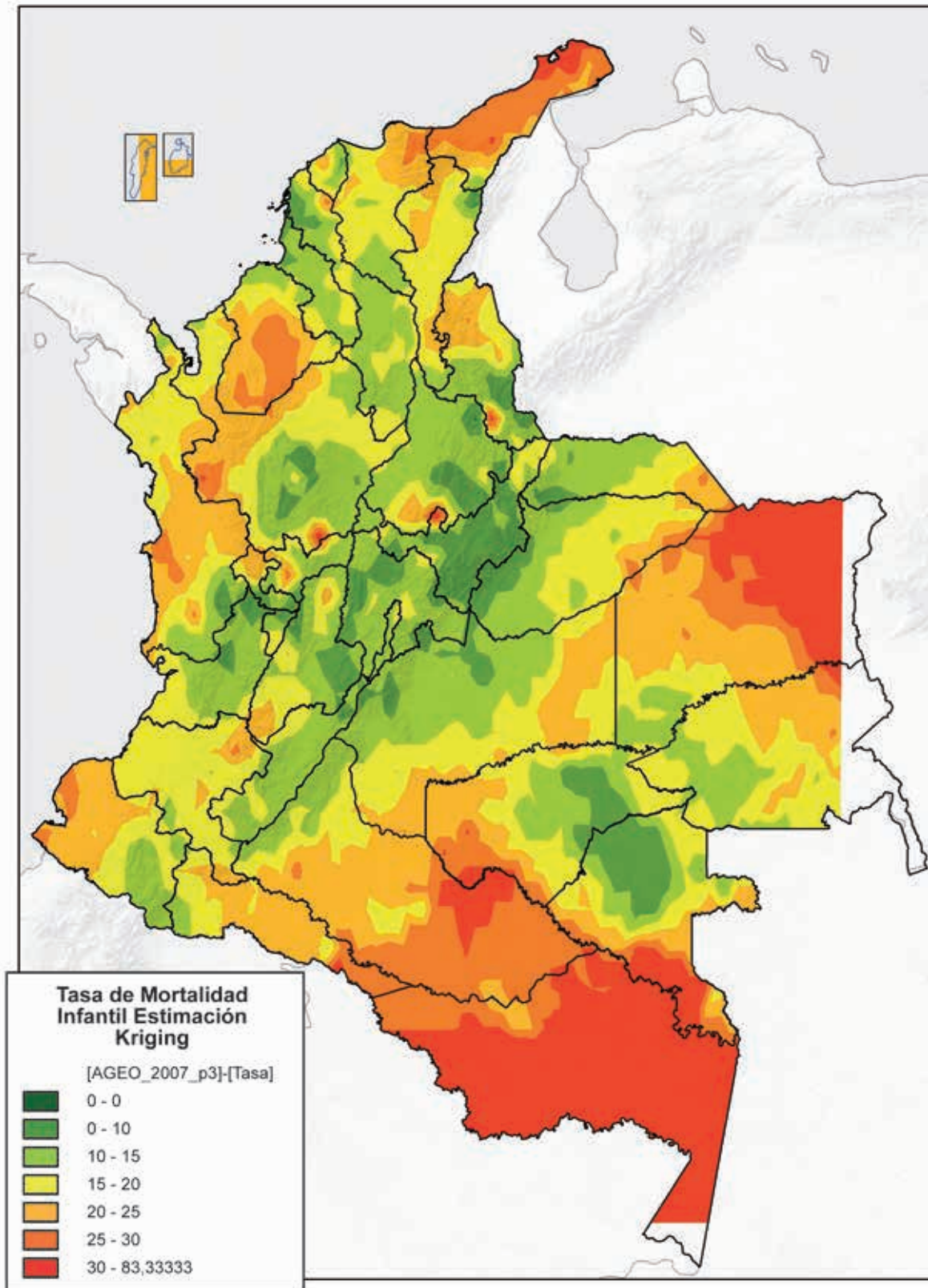
Los mapas de las superficies de estimación permiten reflejar el comportamiento de la TMI de una

Mapa 7. Superficies de estimación de TMI - 2004.



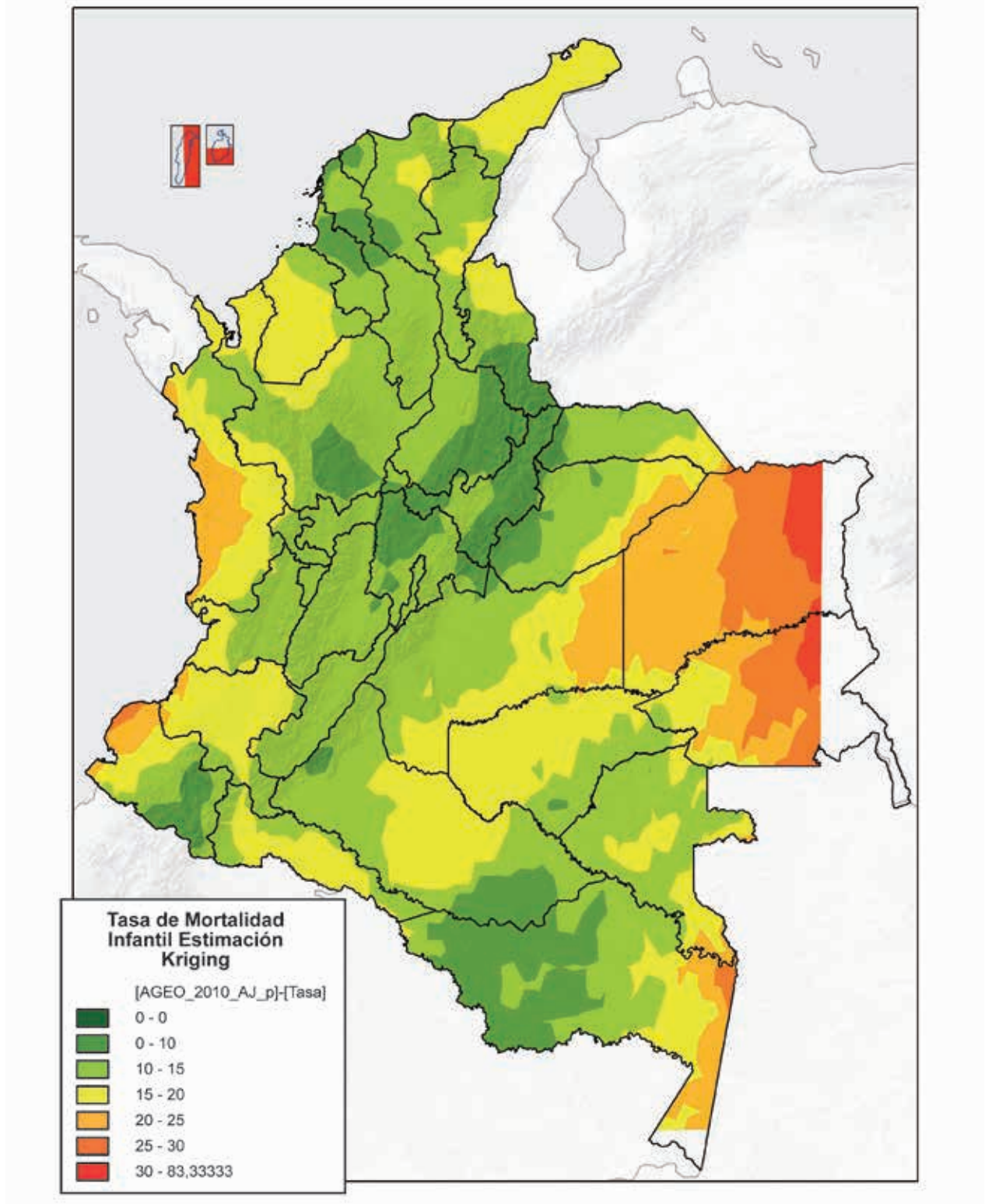
Fuente: elaboración propia.

Mapa 8. Superficies de estimación de TMI - 2007.



Fuente: elaboración propia.

Mapa 9. Superficies de estimación de TMI - 2010.



Fuente: elaboración propia.

manera más clara a través de patrones definidos. Las variaciones (aumentos y descensos) en la misma vigencia o entre periodos, se deben probablemente al cambio en la toma de datos, que se refleja de manera más clara entre 2004 y 2007, donde hubo cambio en la línea base. A su vez, una mejor toma de datos podría explicar un cambio de tipo descendente en el nivel de mortalidad infantil en el país. Sin embargo, los métodos geoestadísticos necesitan que la información de entrada tenga unas condiciones subyacentes para realizar estimaciones con menor grado de incertidumbre, y ante la volatilidad de los datos y la extracción de atípicos, se generan superficies suavizadas

que generalizan el comportamiento de la variable.

Las superficies de los años 2007 y 2010 presentan un corte abrupto en el extremo oriente colombiano, que se debe a la reducción de la zona de influencia determinada por la extracción de valores atípicos y sin información para el proceso, y que no permite la estimación completa para todo el territorio.

Para efectos de verificar los valores de estimación, se establecen comparaciones entre los distintos años en algunos municipios que no tenían dato original de la TMI y en otros que fueron considerados atípicos para los tres años.

Los resultados para estos municipios muestran una gran diferencia entre los valores originales y los estimados, ya que las tasas originales podrían estar sobrestimadas, y con la estimación se generan valores inferiores con respecto a la influencia de los vecinos; además las tasas predichas se mantienen en valores que no fluctúan considerablemente, remarcando la propiedad de generalizar o suavizar que tienen los métodos de interpolación. Los errores de predicción bajan con el pasar de los años, particularmente en 2010, lo que indica una menor variación de los datos, y que conlleva a realizar estimaciones con menor grado de incertidumbre o volatilidad, de acuerdo con la presencia de mejores condiciones requeridas por el estimador.

Tabla 5. Comparación de valores originales y estimados - 2004.

Código	Municipio	TMI original	TMI estimada	Error de predicción
94885	La Guadalupe	ND	21,84	12,36
05475	Murindó	66,66	23,95	11,78
19809	Timbiquí	76,92	19,28	11,78
66572	Pueblo Rico	94,05	17,84	11,72

Fuente: elaboración propia.

Tabla 6. Comparación de valores originales y estimados - 2007.

Código	Municipio	TMI original	TMI estimada	Error de predicción
94885	La Guadalupe	ND	21,92	10,97
05475	Murindó	68,18	23,88	10,77
19809	Timbiquí	82,19	17,70	10,70
66572	Pueblo Rico	68,49	20,46	10,77

Fuente: elaboración propia.

Tabla 7. Comparación de valores originales y estimados - 2010.

Código	Municipio	TMI original	TMI estimada	Error de predicción
94885	La Guadalupe	ND	42,67	4,50
05475	Murindó	58,82	12,48	3,48
19809	Timbiquí	37,04	19,14	3,31
66572	Pueblo Rico	66,95	13,21	2,72

Fuente: elaboración propia.

Conclusiones

El análisis realizado conjuga herramientas estadísticas y espaciales que permiten realizar inferencias sobre el comportamiento y la distribución de la TMI a través del tiempo, a partir de un proceso dividido en etapas, que generan resultados importantes sobre la presencia de patrones de aglomeración de valores altos y bajos en ciertas regiones del país, demostrando que el territorio no solo es un lugar contenedor de información para su representación, sino que el espacio actúa como condicionante en el comportamiento de variables sociodemográficas como la TMI.

Los resultados constituyen un valor agregado fundamental al análisis de la Mortalidad Infantil, confirmando la presencia de focos críticos estadísticamente significativos de la tasa en las zonas periféricas del territorio nacional, y puntos fríos en

la región centro-oriental del país, resaltando, ya sea, la necesidad de mejorar el registro de las EEVV, o, el análisis cuidadoso de los condicionantes que reflejan altos niveles de mortalidad infantil en dichas zonas. Así mismo, las fluctuaciones presentadas por la influencia de la omisión que se exhiben en los nacimientos y las defunciones hacen que se generen patrones cambiantes rápidamente en las regiones de Guaviare, Vaupés, Caquetá y parte de Putumayo y Amazonas, mientras que en la región de la Cordillera de los Andes se observa un patrón de propagación de bajas tasas de mortalidad.

Además, la estimación de valores con precisiones aceptables en sitios no muestreados y en aquellos que presentan un comportamiento atípico es una ventaja en el estudio geoestadístico. Las características

particulares y los posibles condicionantes o causas de las aglomeraciones de valores de la TMI se pueden explicar y modelar a escala local a través de las relaciones espaciales que tienen las variables que explican este fenómeno. Esto nos lleva al planteamiento de un análisis que corresponda al nivel de cada uno de los departamentos con el fin de observar mejor los comportamientos presentados por la tasa a medida que aumenta la calidad de las EEVV.

Con este acercamiento, se logran insumos importantes en el entendimiento del fenómeno de la TMI para Colombia, ya que en los análisis usados para las proyecciones de población se necesita hacer supuestos que sean acordes con las tendencias y realidades presentadas en el país.

Bibliografía

Agua y SIG. (marzo 6 de 2011). Análisis geoestadístico con ArcGIS. Parte 1. Recuperado de <http://www.aguaysig.com/2011/03/modulo-de-geoestadistica-analyst-con.html>

— (marzo 7 de 2011). Análisis geoestadístico con ArcGIS. Parte 2. Recuperado de <http://www.aguaysig.com/2011/03/analisis-geoestadistico-con-arccgis.html>

— (marzo 13 de 2011). Geostatistical Analyst. Análisis geoestadístico con Arcgis. Parte 3. Recuperado de <http://www.aguaysig.com/2011/03/geoestadistical-analyst-analisis.html>

— (mayo 16 de 2012). Cómo escoger un modelo geoestadístico en ArcGIS. Recuperado de <http://www.aguaysig.com/2012/05/como-escooger-un-modelo-geoestadistico.html>

Álvarez, G., Lara, F., Harlow, S. y Denman, C. (2009). Mortalidad infantil y marginación urbana: análisis espacial de su relación en una ciudad de tamaño medio del noroeste mexicano. *Rev Panam Salud Pública*, 26(1): 31-38.

Anselin, L. (1992). *Spatial data analysis with GIS: An introduction to application in the social sciences*. National Center for Geographic In-

formation and Analysis, University of California.

Bodnar, L. (diciembre 6 de 2010). The Use of ArcGIS Geostatistical Analyst Exploratory Spatial Data Analysis and an Integrated Regionalization of Colorado Precipitation and Elevation Data. Recuperado el 21 de enero de 2013, del sitio web de Civil Engineering People: <https://ceprofs.tamu.edu/folivera/TxAGIS/Fall2010/Lacey%20Bodnar.pdf>

Céspedes Panduro, B. (febrero, 2008). Niveles, diferenciales y factores determinantes en la transición de la mortalidad infantil en

el Perú. Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática (INEI). Recuperado de <http://www.inei.gob.pe/biblioineipub/bancopub/Est/Lib0898/Libro.pdf>

Chasco Yrigoyen, C. (mayo de 2010). Sesión 3: Análisis Exploratorio de Datos Espaciales (AEDE): Técnicas básicas. Recuperado el 29 de enero de 2013, del sitio web de la Universidad Autónoma de Madrid: http://www.uam.es/personal_pdi/economicas/coro/courses/Geoespacial/Sesion3_AEDE_Basico.pdf

— (2003). Métodos gráficos del análisis exploratorio de datos espaciales. Recuperado el 29 de enero de 2013, del sitio web de la Asociación Internacional de Economía Aplicada: <http://www.asepelt.org/ficheros/File/Anales/2003%20-%20Almeria/asepeltPDF/93.PDF>.

— Departamento Administrativo Nacional de Estadística, DANE (2009). Estudio de Cabalidad de las Estadísticas Vitales. No publicado.

— (2011). Evaluación de la Cabalidad de las Estadísticas Vitales.

— (2011). Las Estadísticas Vitales en Colombia.

Departamento de Ingeniería Topográfica y Cartografía, Universidad Politécnica de Madrid (s. f.). Variograma. Recuperado el 7 de junio de 2013 de www.geo.upm.es/postgrado/Carlos-Lopez/.../VARIOGRAMA.ppt

Flores Garnica, J., Moreno González, D. (s. f.). Modelaje espacial de la influencia de combustibles forestales sobre la regeneración natural de un bosque perturbado. Recuperado el 22 de mayo de 2013, del sitio web del Colegio de Postgraduados de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación de México: <http://www.colpos.mx/agrocien/Bimestral/2005/may-jun/art-11.pdf>

Gallardo, A. (2006). Geoestadística. Recuperado el 21 de mayo de 2013 de <http://fjferrer.webs.ull.es/Bibliog/Biblio/Geoestadistica.pdf>

Kim, Jay H. (s. f.). Distribution of Infant Mortality and Its Disparity by Socioeconomic Status. Recuperado el 30 de noviembre de 2012, del sitio web de National Center for Health Statistics.

Lentz, J. (2009). Spatial Autocorrelation Statistics. Recuperado el 21 de mayo de 2013, del sitio web de School of the Coast & Environment: <http://www.sce.lsu.edu/cego/>

documents/reviews/geospatial/spatial_autocorrelation.pdf


Mazo, N. (septiembre 16 de 2010). Técnicas y métodos de análisis espacial. Recuperado el 21 de mayo de 2013, de <http://jrubiano.pbworks.com/f/T%C3%A9cnicas+y+M%C3%A9todos+de+An%C3%A1lisis+Espacial+-+Mazo.pdf>

Mitchell, A. (2005). The ESRI guide to GIS Analysis. Vol. 2: Spatial measurements & statistics.

Ponce, E., Fantín, M. Relación entre condiciones habitacionales y mortalidad infantil en la provincial del Chaco (Argentina). Determinación de diferentes escenarios posibles. Recuperado el 11 de abril de 2013, del sitio web de National Center for Health Statistics.

Portland State University. Exploratory Spatial Data Analysis (ESDA). Recuperado el 29 de enero de 2013, de <http://web.pdx.edu/~jduh/courses/geog492w10/Week8b.pdf>

Smith, M., Goodchild, M. Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and Software Tools. Recuperado el 26 de agosto de 2013, de <http://www.spatialanalysisonline.com/output/index.html>



Ajuste de estimadores mediante la técnica de propensity score en encuestas complejas

Hugo Andrés Gutiérrez Rojas
Luisa Fernanda Rojas Niño

Ajuste de estimadores mediante la técnica de propensity score en encuestas complejas

Hugo Andrés Gutiérrez Rojas: Estadístico, magíster y estudiante del doctorado de Estadística de la Universidad Nacional de Colombia. Decano de la Facultad de Estadística de la Universidad Santo Tomás. Correo electrónico: hugogutierrez@usantotomas.edu.co

Luisa Fernanda Rojas Niño: Estadística de la Universidad Santo Tomás. Ejecutiva estadística junior de Nielsen de Colombia. Correo electrónico: luisa.rojas@nielsen.com

Fecha de recepción: 20 de julio de 2013
Fecha de aceptación: 1 de noviembre de 2013

Resumen: para enfrentar el sesgo que genera la ausencia de respuesta en los estimadores de totales de interés, es posible ajustar su forma funcional mediante la técnica de *propensity score* (PS). En este artículo se realizaron simulaciones de Monte Carlo bajo un esquema de muestreo proporcional al tamaño sin reemplazo (PIPT) para comparar el desempeño de algunos estimadores ajustados mediante PS. Además, se utilizaron estos estimadores en la Gran Encuesta Integrada de Hogares (GEIH) y se compararon los resultados obtenidos, cotejando cuál es la mejor opción en términos de sesgo y eficiencia estadística.

Palabras clave: ausencia de respuesta, *propensity score*, simulaciones de Monte Carlo.

Abstract: the aim of this paper is to review some adjusted estimators of the population total by means of the Propensity Score (PS) technique to address nonresponse. In order to evaluate the performance of this approach, Monte Carlo simulations were implemented under without replacement proportional to size sampling (PIPS). We also used the proposed estimators in the Great Integrated Household Survey (GEIH) and compared the estimators in terms of bias and efficiency.

Key words: no response, Propensity Score, Monte Carlo simulations.

1. Introducción

Al analizar datos provenientes de encuestas, en muy pocas ocasiones es posible disponer de un archivo o base de datos completa. La presencia de datos faltantes debido a la ausencia de respuesta total o parcial es una situación recurrente en el desarrollo de una encuesta. Ignorar este fenómeno puede generar problemas en la fase de inferencia estadística, añadiendo sesgos a los estimadores de los parámetros de interés.

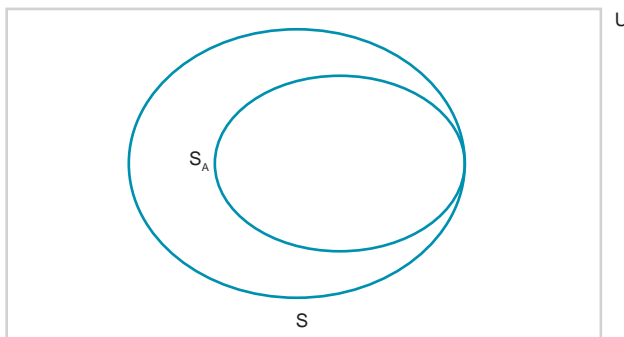
El investigador puede encontrar paquetes estadísticos que, mediante alguna rutina particular, generan sus inferencias al asumir que la base de datos está completa, y las opciones que incorporan para hacer frente a la ausencia de respuesta no siempre son advertidas por el usuario. En la literatura estadística está ampliamente documentado que la mejor forma de

enfrentar la ausencia de respuesta es evitándola mediante la correcta planeación en el diseño de la encuesta (Lohr, 2010). Sin embargo, cuando se presenta este fenómeno y se hace uso de procedimientos estadísticos para sustituir la información faltante o estimarla, restándole importancia a que las unidades de observación que no responden pueden diferir en forma importante de las que sí lo hacen, entonces el proceso de estimación induciría sesgos importantes.

Ante la ausencia de respuesta, es posible realizar un proceso de inferencia en el subconjunto de individuos inducido por los elementos que no respondieron, haciendo uso de la metodología de *propensity score* (Rosenbaum & Rubin, 1983), cuya aplicación en los estudios observacionales hace que disminuya la aparición de sesgos, cuando hay ausencia de respuesta, para así acercarse en precisión y fiabilidad a los estimadores ajustados.

(Kim & Riddles, 2012). Después de esta breve introducción, la sección 2 muestra algunos conceptos básicos teniendo en cuenta en primer lugar los principios elementales del muestreo, y en segundo lugar una explicación detallada de la ausencia de respuesta. La sección 3 incluye la definición teórica de algunos estimadores de muestreo ajustados por el *propensity score* para enfrentar la ausencia de respuesta. La cuarta sección evidencia las simulaciones de Monte Carlo que se realizaron para calcular el sesgo, el error cuadrático medio (ECM) y la eficiencia para un esquema de muestreo sin reemplazo proporcional al tamaño (PIPT), teniendo en cuenta el tamaño de la muestra, el porcentaje de respondientes y la correlación entre las variables auxiliares y la variable de interés. La sección 5 muestra la aplicación real de esta técnica en la Gran Encuesta Integrada de Hogares (GEIH). Por último, la sección 6 discute los hallazgos de este trabajo de investigación.

Figura 1. Esquema general de la muestra de respondientes S_A , contenida en la muestra S , a su vez seleccionada de la población U .



Fuente: elaboración propia.

La figura 1 explica este proceso de expansión en dos fases. S representa la muestra seleccionada y S_A , la muestra de los respondientes efectivos cuyo complemento representa el conjunto de los no respondientes que generan los datos faltantes en la base de datos final. El espíritu de la metodología que se expone en este artículo induce un

proceso de expansión de la muestra de respondientes a la muestra original, mediante la técnica de *propensity score* que, a su vez, se expande al universo mediante los factores de expansión inducidos por el diseño de muestreo original, que la mayoría de las veces resulta ser complejo, más aún en el sector de las estadísticas oficiales

2. Conceptos fundamentales

2.1. Principios de muestreo

Suponga una población de N elementos definidos por un conjunto de rótulos denotados por $U = \{1, 2, \dots, N\}$. Note que N es el tamaño de la población, no siempre conocido de antemano. Asumiendo que se selecciona una muestra aleatoria S de la población, cuyo tamaño de muestra es $n(s)$, entonces se define la probabilidad de que un individuo sea incluido en la muestra, denotada como:

$$\pi_k = Pr(k \in S) = \sum_{S \in Q} p(S)$$

En donde Q se conoce como el soporte de muestreo. Suponiendo que el objetivo principal de la realización de una encuesta, basada en

una muestra probabilística, es conocer el total de una característica de interés dado por:

$$t_y = \sum_{k \in U} y_k$$

Entonces, se define el estimador de Horvitz-Thompson (HT) como:

$$\hat{t}_{y,\pi} = \sum_S \frac{y_k}{\pi_k} = \sum_S d_k y_k \quad (1)$$

Nótese que si todas las probabilidades de inclusión son mayores a cero, este estimador es insesgado para el total poblacional. El estimador HT tiene varianza dada por:

$$Var(\hat{t}_{y,\pi}) = \sum_U \Delta_{kl} \frac{y_k y_l}{\pi_k \pi_l}$$

Y la estimación insesgada de la anterior expresión es:

$$\widehat{Var}(\hat{t}_{y,\pi}) = \sum_S \frac{\Delta_{kl} y_k y_l}{\pi_{kl} \pi_k \pi_l}$$

En donde π_{kl} es conocida como la probabilidad de inclusión de segundo orden y $\Delta_{kl} = (\pi_{kl} - \pi_k \pi_l)$. Dado que el estimador HT no utiliza información secundaria, es posible definir un nuevo estimador de mejor eficiencia al incluir información auxiliar que se encuentre bien relacionada (linealmente) con la característica de interés y_k , mediante el siguiente modelo lineal:

$$y_k = \mathbf{x}'_k \mathbf{B} + \varepsilon_k$$

En donde $\mathbf{x}_k = (x_{1k}, \dots, x_{pk})'$ es el vector de variables auxiliares. Además, $E_\xi(\varepsilon_k) = 0$ y $Var_\xi(\varepsilon_k) = c_k \sigma^2$. Por lo tanto,

$$E_\xi(Y_k) = \mathbf{X}'_k \mathbf{B}$$

$$Var_\xi(Y_k) = c_k \sigma^2$$

Nótese que ε_k son variables aleatorias independientes e idénticamente distribuidas y C_k es un valor que

determina la naturaleza homoscedástica o heteroscedástica del modelo. Bajo esta perspectiva, es un vector de coeficientes de regresión. En virtud de lo anterior, se define el estimador de regresión general dado por la siguiente expresión:

$$\hat{t}_{y,greg} = \sum_U \mathbf{X}'_k \widehat{\mathbf{B}} + \sum_S \frac{y_k - \mathbf{x}'_k \widehat{\mathbf{B}}}{\pi_k} \quad (2)$$

La varianza aproximada y la estimación de la varianza de este estimador es la siguiente:

$$AVar(\hat{t}_{y,cal}) = \sum_U \sum_U \Delta_{kl} \frac{E_k E_l}{\pi_k \pi_l}$$

$$\widehat{Var}(\hat{t}_{y,cal}) = \sum_S \sum_S \frac{\Delta_{kl} e_k e_l}{\pi_{kl} \pi_k \pi_l}$$

respectivamente. Siendo,

$$E_k = y_k - \mathbf{x}'_k \mathbf{B} \text{ y } e_k = y_k - \mathbf{x}'_k \widehat{\mathbf{B}}$$

En donde,

$$\mathbf{B} = \left(\sum_U \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k} \right)^{-1} \left(\sum_U \mathbf{x}_k y_k \right)$$

$$\widehat{\mathbf{B}} = \left(\sum_S \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k \pi_k} \right)^{-1} \left(\sum_S \frac{\mathbf{x}_k y_k}{c_k \pi_k} \right)$$

Ahora, suponiendo que se conoce el vector de totales poblacionales de las variables auxiliares, dado por:

$$\mathbf{t}_X = \sum_{k \in U} \mathbf{x}'_k = (t_{x1}, t_{x2}, \dots, t_{xp})'$$

Entonces, un estimador aproximadamente insesgado para el total poblacional t_y , denotado como estimador de calibración, se define como:

$$\hat{t}_{y,cal} = \sum_{k \in S} w_k y_k \quad (3)$$

En donde w_1, w_2, \dots, w_n , constituye un conjunto de ponderaciones que minimizan una pseudo-distancia entre \mathbf{W}_k y el inverso de la probabilidad de inclusión $d_k = 1/\pi_k$, tales que respeten la siguiente ecuación de calibración:

$$\hat{\mathbf{t}}_{X,cal} = \sum_S w_k \mathbf{x}_k = \mathbf{t}_X$$

Es decir, al ponderar las variables auxiliares en la muestra con estos nuevos pesos W_k , se reproduce sin error cada total auxiliar. Es posible demostrar que la varianza aproximada y la estimación de la varianza de este estimador de calibración es asintóticamente equivalente a la varianza estimada del estimador de regresión. Algunas ventajas al utilizar un estimador de calibración son consistencia, eficiencia y robustez (Gutiérrez, 2009).

2.2. Ausencia de respuesta

Para poder determinar los parámetros de interés en una población particular, los investigadores hacen uso de las encuestas. Sin embargo, las encuestas adolecen con regularidad de datos incompletos, ya sea porque faltan respuestas de unidades enteras (es decir, cuando las personas consideradas en el estudio no desean participar o cuando no es posible ponerse en contacto con ellas mediante el mecanismo de selección establecido) o porque faltan respuestas aisladas (es decir, cuando algunas personas no contestan determinadas preguntas del cuestionario). La ausencia de respuesta en unidades completas se suele compensar¹ mediante el ajuste de los pesos muestrales (que bajo este contexto son $d_k = 1/\pi_k \circ w_k$).

La finalidad de estas herramientas es analizar los datos de forma completa para minimizar las consecuencias de la falta de respuesta. La precisión de los estimadores se ve afectada por la reducción del número total de casos, o de uno o más elementos, en tanto que la validez se verá afectada debido a los posibles sesgos generados por este fenómeno. Los datos faltantes dan a lugar a una pérdida considerable del tamaño de la muestra. En las encuestas pequeñas, la mayoría de veces se ignora la ausencia de respuesta, y se hace la inferencia a partir del conjunto final

¹ Cabe resaltar que el diseño y formulación de la encuesta debe estar bien sustentado para que los datos faltantes sean mínimos. Además se recomienda realizar una submuestra representativa de quienes no responden e inferir acerca de estos, y por último realizar un modelo predictivo de quienes no contestaron, para estimar los datos faltantes.

de datos, que rara vez representa la población objetivo. Lohr (2010) menciona que el principal problema de la ausencia de respuesta es el sesgo potencial de los estimadores, debido a que, por ejemplo, si la media de la población en el estrato de las personas que no responden difiere de la media en el estrato de los que sí, el estimador de la media de la población utilizando solo a quienes responden tendrá un sesgo considerable.

El tratamiento de posibles datos perdidos debe considerarse fundamentalmente en tres momentos del desarrollo de todo estudio de investigación: a) durante la etapa de diseño, mediante la adición al tamaño muestral mínimo requerido de una proporción de elementos que permita compensar las posibles pérdidas de datos; b) durante la etapa de recolección de datos, mediante un adecuado monitoreo de la calidad de los datos que permita recuperar datos perdidos; y c) en la etapa de procesamiento de información que permitirá valorar la medida en que se ven afectadas las estimaciones por la pérdida de datos.

Suponiendo que φ_k corresponde a la probabilidad de respuesta de un individuo, Y_k es la característica de interés y $z_k = (z_{1k}, z_{2k}, \dots, z_{pk})'$ es un vector de información auxiliar conocida para cada unidad de la muestra, entonces podemos distinguir los siguientes tipos de datos faltantes (Lohr, 2010):

- Completamente al azar: si φ_k no depende de Z_k ni de Y_k y si la submuestra de personas que sí respondieron es representativa de la muestra seleccionada originalmente. Si los φ_k son todos iguales y los eventos son condicionalmente dependientes entre sí e inde-

pendientes del proceso de selección, entonces se dice que los datos faltantes son MCAR (Missing Completely at random).

- Al azar: si φ_k depende de Z_k pero no de Y_k , se considera que los datos son MAR (Missing at random). En este caso, un modelo estocástico podría explicar el mecanismo de ausencia de respuesta.
- No aleatoria: si φ_k depende de Z_k y no Y_k puede explicarse por completo mediante los valores de Y_k , se dice que los datos faltantes son NMAR (Not missing at random).

La mayoría de las veces no se conoce qué tipo de ausencia de respuesta está presente, en un problema de investigación. Aunque, si los coeficientes de un modelo logístico son significativamente diferentes de cero, es probable que los datos faltantes no sean MCAR (Lohr, 2010).

3. Propensity score para enfrentar la ausencia de respuesta

Rosenbaum & Rubin (1983) propusieron la estimación a través del *propensity score*, como una forma de reducir el sesgo en la estimación del efecto de un tratamiento sobre un conjunto de datos observados. En términos de una muestra probabilística, el *propensity score* indicaría la probabilidad de que el elemento k-ésimo de la muestra haya sido un respondiente efectivo, teniendo en cuenta sus covariables. En este sentido el *propensity score* puede estimarse² con un modelo de regresión logística, utilizando como variables explicativas diferentes características individuales que determinan

la posibilidad de que una persona conteste o no lo haga.

Para efectos de notación, se definen las siguientes expresiones: S es la muestra completa, seleccionada mediante un diseño muestral $p(s)$, S_A es la muestra de los elementos que efectivamente contestaron la encuesta, S_{NA} es la muestra de los elementos que no contestaron, habiendo sido incluidos en la muestra S , Y_k es la característica de interés en la k-ésima unidad de muestreo, y por último π_k es la probabilidad de inclusión de cada elemento. Siendo R_k una variable aleatoria que define la respuesta (o la no respuesta) del elemento k, se expresa como:

$$R_k = \begin{cases} 1, & \text{si el elemento responde} \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto, se conoce el valor de Y_k si y solamente si $R_k = 1$. Dado lo anterior, la probabilidad de que el elemento k responda (*propensity score*) está dada por:

$$\varphi_k = Pr(R_k = 1) = E(R_k) = Pr(k \in S_A)$$

Se debe tener en cuenta que para estimar φ_k , es posible recurrir a un modelo logístico, en donde:

$$\hat{\varphi}_k = \frac{e^{z_k \hat{\beta}}}{1 + e^{z_k \hat{\beta}}}$$

Siendo $\hat{\beta}$ un vector de estimaciones de coeficientes de regresión. Bajo este escenario, es posible realizar una corrección a los factores de expansión haciendo uso de las estimaciones del *propensity score*. Por lo tanto, el producto $\pi_k \varphi_k$ expande la característica de interés del respondiente efectivo a toda la población, puesto que φ_k expande de S_A a S y expande de S a U . Por lo tanto, se define el *estimador HT ajustado* como:

$$\hat{t}_{y,HT}^a = \sum_{S_A} \frac{y_k}{\pi_k \hat{\varphi}_k} \quad (4)$$

El anterior estimador es insesgado respecto al diseño de muestreo. Sin embargo, es posible definir otro tipo

² Existen desarrollos importantes en términos de estimación e inferencia en este tipo de modelos. El lector interesado puede encontrar en Guo & Fraser (2009) una fuente valiosa de información acerca de esta metodología.

de estimadores, más eficientes que el HT ajustado. De esta forma el *estimador de calibración ajustado a nivel de la muestra* se define como:

$$\hat{t}_{y,cal}^a = \sum_{S_A} \frac{y_k}{\pi_k \hat{p}_k} \quad (5)$$

En donde \hat{P}_k es una estimación calibrada de que satisface las siguientes restricciones de calibración.

$$\hat{t}_{z,cal}^a = \sum_{S_A} \frac{z'_k}{\pi_k \hat{p}_k} = \sum_S \frac{z'_k}{\pi_k} = \hat{t}_{z,\pi}$$

También es posible definir, siguiendo el mismo espíritu del anterior enfoque, el *estimador de calibración ajustado a nivel de la población*, teniendo en cuenta que se tiene acceso al total poblacional de las variables auxiliares t_z . Este nuevo estimador se define como:

$$\hat{t}_{y,cal2}^a = \sum_{S_A} \frac{y_k}{\pi_k \hat{p}_k} \quad (6)$$

En donde satisface las siguientes restricciones de calibración:

$$\hat{t}_{z,cal2}^a = \sum_{S_A} \frac{z'_k}{\pi_k \hat{p}_k} = \sum_U z_k = t_z$$

Por último, se define el *estimador GREG ajustado* que mejora la eficiencia del estimador HT ajustado al incluir información auxiliar, suponiendo que las variables auxiliares están relacionadas con la característica de interés mediante un modelo de regresión lineal y está definido como:

$$\hat{t}_{y,greg}^a = \sum_S \frac{y_k}{\pi_k} + \sum_{S_A} \frac{y_k - \hat{y}_k}{\pi_k \hat{\phi}_k} = \sum_S \frac{\hat{B}_0 + \hat{B}_1 z_k}{\pi_k} + \sum_{S_A} \frac{y_k - \hat{B}_0 - \hat{B}_1 z_k}{\pi_k \hat{\phi}_k} \quad (7)$$

En donde \hat{B}_0 y \hat{B}_1 son estimadores de B_0 y B_1 , respectivamente, inducidos por un modelo de regresión de la forma $y_k = B_0 + B_1 z_k + \varepsilon_k$

4. Simulaciones

En esta sección se realizan varias simulaciones³ mediante el método de Monte Carlo sobre una población de tamaño $N=10000$ para comparar el desempeño de los estimadores de Horvitz-Thompson ajustado (4), el estimador de calibración ajustado a nivel de muestra (5), el estimador de calibración ajustado a nivel de población (6) y por último el estimador de regresión ajustado (7).

En esta simulación, se define f como la fracción de muestreo y $P(S_A)$ como el porcentaje de unidades que responden (o tasa de respuesta). Para cada estimador se evalúa el sesgo, error cuadrático medio y la eficiencia en cada uno de los escenarios propuestos. La eficiencia⁴ se evalúa con respecto al estimador de Horvitz-Thompson ajustado.

Las variables auxiliares y de interés se simularon de acuerdo con el siguiente esquema: en primer lugar el número de simulaciones en cada escenario se fijó en 10000, la creación de la variable auxiliar fue inducida por una distribución normal con media 50 y varianza 4. La característica de interés se simuló teniendo en cuenta un modelo de regresión, para el cual el error tuvo media cero y su varianza se modificó para poder establecer escenarios dependientes de la correlación entre X , Y . Se evaluaron escenarios de correlación de -1 a 1. Por lo anterior, la característica de interés está dada por:

$$y_k = B_0 + B_1 z_k + \varepsilon_k$$

Donde $\beta_0 = 1000$ y $\beta_1 = 0.4$ para lograr escenarios de correlación positiva y $\beta_1 = -0.4$ para lograr escenarios de correlación negativa. Por otro lado, la probabilidad de respuesta de un individuo estuvo determinada por el siguiente modelo de regresión logística.

$$\phi_k = \Pr(R_k = 1) = \frac{e^{x'_k \mathbf{B}}}{1 + e^{x'_k \mathbf{B}}}$$

Siendo $\mathbf{B} = (B_1, B_2)'$ y los valores de B_1 y B_2 se modificaron para lograr tasas de respuesta del 1%, 50% y 99%. El diseño de muestreo considerado es un muestreo sin reposición proporcional a los valores de la variable auxiliar X (PIPT).

4.1. Estimador de Horvitz & Thompson ajustado

En la tabla 1 correspondiente al sesgo del estimador de Horvitz & Thompson en PIPT, se observa que cuando la tasa de respuesta es del 1% existe un sesgo considerable. También se evidencia que al pasar de tasas de respuesta 1% al 50%, se registra una disminución importante del sesgo. Por último, se observa que no existe un patrón de comportamiento en la variabilidad de la correlación.

De la tabla 2, se evidencia que el ECM se comporta muy semejante al sesgo en cuanto a las tasas de respuesta. En lo que corresponde al tamaño de muestra se evidencia una conducta clara en la disminución del ECM para tasas de respuesta del 99%. No se muestra un patrón de comportamiento específico con respecto a la correlación.

4.2. Estimador de calibración ajustado

En la tabla 3, se observa que el estimador de calibración ajustado es aproximadamente insesgado; aunque para tasas de respuesta del 1%, el sesgo es mayor en comparación con los demás porcentajes de respondientes. También se evidencia que al pasar del 10% al 50% en el tamaño de la muestra, se registra una disminución del sesgo en todos los porcentajes de respondientes. En última medida, se observa que no existe un patrón de comportamiento específico con respecto a la correlación.

De la tabla 4, se concluye que el ECM del estimador de calibración

³ Los códigos computacionales en el lenguaje R se proveen mediante comunicación con el primer autor.

⁴ Las cifras del ECM están dadas en miles de millones.

Tabla 1. Sesgo del estimador de Horvitz & Thompson.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S _A)									
Cor -0.9	-68.71	-3.10	-0.07	-48.15	-0.93	0.01	-35.95	0.04	-0.02
Cor -0.7	-69.55	5.66	-0.01	-57.09	-0.77	-0.01	-22.76	-0.16	0.01
Cor -0.5	-65.35	-0.24	-0.03	-42.70	-0.32	0.01	36.99	-0.33	0.00
Cor -0.3	-68.10	-2.37	-0.03	-53.97	1.91	0.01	-27.26	0.23	0.00
Cor -0.1	-71.65	-0.23	-0.04	-59.77	-2.04	-0.03	-45.77	0.11	-0.02
Cor 0.1	-67.53	-2.31	-0.03	-53.98	-0.50	0.00	-39.40	1.26	0.01
Cor 0.3	-68.34	-2.23	-0.03	-48.80	-1.11	-0.03	-30.10	-0.55	0.01
Cor 0.5	-68.09	-2.32	-0.03	-38.09	-0.97	-0.01	-43.62	-0.19	0.00
Cor 0.7	-70.53	-2.85	-0.07	-56.30	-1.46	0.00	38.90	-0.57	-0.02
Cor 0.9	-71.26	-2.96	0.02	-57.99	-1.85	0.00	-27.16	2.95	-0.01

Fuente: elaboración propia.

Tabla 2. ECM del estimador de Horvitz & Thompson.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S _A)									
Cor -0.9	112369.60	5041.88	3.62	763567.50	8640.65	4.38	781516.90	4822.86	0.80
Cor -0.7	123094.20	111315.70	2.20	138241.60	6193.29	1.82	1156335.00	4051.25	1.95
Cor -0.5	132109.70	18511.87	4.93	1023893.00	25607.77	1.58	55004010.00	8766.36	0.27
Cor -0.3	86603.42	4812.76	1.23	249593.60	27508.95	1.25	1001443.00	5774.08	0.48
Cor -0.1	69522.66	17800.74	4.72	107720.90	4720.77	1.73	181409.10	6122.19	1.61
Cor 0.1	136190.30	11561.79	3.10	99618.14	11949.36	1.11	240478.30	23650.65	1.27
Cor 0.3	97541.00	9715.53	2.27	325580.70	4745.29	1.16	1000428.00	5439.95	0.84
Cor 0.5	134746.60	6614.66	3.28	2172195.00	10236.09	1.61	191287.60	5236.28	0.60
Cor 0.7	84969.52	8083.46	3.21	224171.40	7927.85	1.54	59793940.00	3110.86	1.81
Cor 0.9	101377.90	6760.31	9.32	284342.20	4311.79	1.23	2798114.00	59660.89	0.49

Fuente: elaboración propia.

Tabla 3. Sesgo del estimador de calibración ajustado a nivel de la muestra.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S _A)									
Cor -0.9	1.29	-0.36	-0.31	0.25	-0.15	-0.04	0.17	0.15	-0.07
Cor -0.7	-0.29	0.02	-0.06	-0.84	-0.16	0.06	-0.37	0.15	0.00
Cor -0.5	-0.54	0.65	-0.35	0.39	-0.11	-0.06	0.34	-0.02	-0.01
Cor -0.3	0.65	-0.32	0.08	0.70	-0.01	0.16	-0.18	0.08	-0.08
Cor -0.1	0.09	0.01	0.07	0.59	0.23	0.24	0.18	0.05	0.05
Cor 0.1	-0.59	0.31	0.13	-0.13	-0.27	0.09	-0.47	-0.02	-0.08
Cor 0.3	1.66	-0.20	0.35	-0.94	-0.15	-0.04	0.04	-0.19	-0.03
Cor 0.5	-0.59	0.10	0.28	-0.31	-0.20	0.13	-0.12	0.00	0.06
Cor 0.7	-0.54	0.57	0.26	-0.39	0.00	-0.09	0.05	0.00	0.14
Cor 0.9	-0.16	-0.01	-0.45	-0.06	-0.04	0.06	0.29	0.03	0.06

Fuente: elaboración propia.

ajustado a nivel de la muestra se se comporta muy semejante a su sesgo, y se puede deducir que efectivamente disminuye a medida que se aumenta el tamaño de muestra y el porcentaje de respondientes. No se evidencia un patrón definido en cuanto a la correlación.

En la tabla 5, se muestra la eficiencia del estimador de calibración ajustado a nivel de la muestra con respecto al estimador HT ajustado. Se observa una mayor eficiencia del estimador mayor de calibración. A medida que se aumenta el porcentaje de respondientes aumenta el tamaño de muestra y a medida que el tamaño de muestra aumenta, también lo hace la eficiencia, debido a que el estimador definido de HT ajustado disminuye su varianza en estos escenarios. Con respecto a la correlación no se evidencia ningún patrón.

Los resultados del estimador de calibración ajustado de nivel de la población son idénticos a los encontrados con el estimador de calibración ajustado a nivel de la muestra y se omiten, puesto que bajo un diseño de muestreo PIPT, el estimador de Horvitz-Thompson reproduce exactamente al total poblacional. Es decir:

$$\hat{t}_{x,\pi} = \sum_s \frac{x_k}{\pi_k} = \sum_s \frac{x_k}{\left(\frac{nx_k}{t_x}\right)} = t_x$$

4.3. Estimador GREG ajustado

En la tabla 6, se evidencia el sesgo del estimador GREG ajustado, observando que todos los valores del sesgo son despreciables. Para tasas de respuesta del 1% el sesgo es ligeramente mayor en comparación con los demás escenarios.

En la tabla 7 se observa el ECM del estimador GREG ajustado. Se observa que en general, no existe un ECM alto. Sin embargo, se nota que el estimador tiene más variabilidad cuando disminuye la tasa de

Tabla 4. ECM del estimador de calibración ajustado a nivel de la muestra.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S_A)	1%	50%	99%	1%	50%	99%	1%	50%	99%
Cor -0.9	302.89	20.39	0.95	256.08	10.34	0.04	176.30	6.80	0.06
Cor -0.7	249.37	20.74	0.06	192.45	12.39	0.05	198.69	7.55	0.01
Cor -0.5	293.93	23.49	1.24	235.01	12.60	0.04	207.74	9.51	0.00
Cor -0.3	296.02	16.60	0.08	259.73	18.04	0.24	192.78	11.22	0.06
Cor -0.1	317.96	17.44	0.07	253.12	10.12	0.56	193.18	10.70	0.03
Cor 0.1	287.20	21.15	0.23	270.72	12.71	0.11	217.84	8.30	0.08
Cor 0.3	318.01	18.42	1.31	268.71	15.77	0.02	214.63	6.74	0.02
Cor 0.5	342.58	18.14	0.84	257.92	13.02	0.20	240.19	5.90	0.04
Cor 0.7	283.85	25.35	0.76	211.68	11.57	0.12	239.71	9.16	0.21
Cor 0.9	268.98	26.04	2.18	258.79	14.09	0.05	243.22	10.40	0.05

Fuente: elaboración propia.

Tabla 5. Eficiencia del estimador de calibración ajustado a nivel de la muestra.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S_A)	1%	50%	99%	1%	50%	99%	1%	50%	99%
Cor -0.9	370.99	247.23	3.82	2981.77	835.73	122.32	4433.00	709.50	13.97
Cor -0.7	493.63	5367.50	37.54	718.33	499.88	38.98	5819.86	536.59	277.11
Cor -0.5	449.45	788.06	3.96	4356.78	2033.10	37.28	264779.70	921.36	59.68
Cor -0.3	292.56	289.89	16.00	960.96	1525.14	5.16	5194.63	514.54	7.79
Cor -0.1	218.65	1020.65	64.52	425.57	466.54	3.09	939.07	572.12	48.04
Cor 0.1	474.21	546.58	13.21	367.98	940.03	10.10	1103.93	2849.85	15.56
Cor 0.3	306.72	527.49	1.73	1211.63	300.97	49.50	4661.27	806.68	45.84
Cor 0.5	393.33	364.67	3.90	8422.05	785.89	8.23	796.41	888.08	14.39
Cor 0.7	299.35	318.92	4.21	1059.01	685.21	13.19	249443.60	339.47	8.83
Cor 0.9	376.90	259.61	4.27	1098.75	306.09	25.84	11504.50	5734.60	10.47

Fuente: elaboración propia.

Tabla 6. Sesgo del estimador GREG ajustado.

f	10%			20%			50%		
	1%	50%	99%	1%	50%	99%	1%	50%	99%
P(S_A)	1%	50%	99%	1%	50%	99%	1%	50%	99%
Cor -0.9	0.30	0.02	0.00	0.24	0.00	0.01	0.19	0.00	0.00
Cor -0.7	0.33	-0.04	0.01	0.28	0.01	0.01	0.12	0.01	0.00
Cor -0.5	0.29	0.00	0.01	0.13	0.01	0.00	-0.07	0.00	0.00
Cor -0.3	0.30	0.02	-0.01	0.26	0.02	0.00	0.17	0.00	0.00
Cor -0.1	0.37	0.00	0.01	0.26	0.02	0.00	0.36	0.00	0.00
Cor 0.1	-0.18	0.00	0.00	-0.23	0.01	0.00	-0.22	-0.01	0.00
Cor 0.3	-0.26	0.00	0.01	-0.26	-0.01	0.00	-0.16	0.01	0.00
Cor 0.5	-0.30	-0.01	0.01	-0.15	-0.01	0.01	-0.21	0.00	0.00
Cor 0.7	-0.30	-0.01	-0.01	-0.26	-0.01	0.00	0.19	-0.01	0.00
Cor 0.9	-0.32	-0.02	0.01	-0.26	0.00	0.01	-0.15	0.01	0.00

Fuente: elaboración propia.

respuesta. En los escenarios en donde la tasa de respuesta es del 50% y 99%, el ECM del estimador es estable y bajo. Por último, se observa que no existe un patrón de comportamiento en la correlación.

En la tabla 8, se observa la eficiencia del estimador GREG ajustado. A medida que se aumenta el porcentaje de respondientes disminuye la eficiencia. El comportamiento del estimador con respecto a la correlación no sigue un patrón definido.

5. Aplicación real

En esta sección, se realiza la aplicación de esta metodología a la base de datos de la Gran Encuesta Integrada de Hogares (GEIH) del mes de diciembre de 2012, que se encuentra disponible en la página web del DANE. En esta encuesta se solicita información múltiple, como, por ejemplo: condiciones de empleo de los respondientes, además de características generales de la población como género, edad, estado civil y nivel educativo. Entre otras, se pregunta sobre fuentes de ingresos y gastos. La GEIH proporciona estadísticas esenciales relacionadas con la situación demográfica, social y económica de la población colombiana.

Para implementar la metodología expuesta en este artículo, se consideran en particular las siguientes variables auxiliares:

- Género (GEN): correspondiente a las variables explicativas X_1 y X_2

$$X_{1k} = \begin{cases} 1, & \text{si el elemento } k \text{ es mujer} \\ 0, & \text{en otro caso.} \end{cases}$$

$$X_{2k} = 1 - X_{1k}$$

- Escolaridad (ESC): es la variable explicativa X_3 , la cual es el número de años cursados y aprobados.

Tabla 7. ECM del estimador GREG ajustado.

f	10%			20%			50%		
P(S_A)	1%	50%	99%	1%	50%	99%	1%	50%	99%
Cor -0.9	2.43	0.64	0.57	12.17	0.43	0.24	19.13	0.18	0.06
Cor -0.7	2.65	5.80	0.56	2.21	0.46	0.24	27.68	0.14	0.06
Cor -0.5	2.24	1.12	0.55	127.21	0.72	0.24	835.41	0.42	0.06
Cor -0.3	3.21	0.83	0.57	10.42	0.83	0.25	24.99	0.35	0.06
Cor -0.1	9.87	4.48	0.62	11.06	1.07	0.27	14.81	1.63	0.06
Cor 0.1	23.10	2.40	0.60	16.72	1.66	0.25	13.55	0.59	0.07
Cor 0.3	3.60	0.90	0.60	16.13	0.42	0.26	54.60	0.48	0.06
Cor 0.5	2.80	0.72	0.57	102.02	0.38	0.27	3.96	0.22	0.06
Cor 0.7	2.26	0.79	0.58	3.45	0.40	0.24	1396.73	0.13	0.06
Cor 0.9	2.29	0.72	0.56	5.78	0.32	0.26	62.31	1.46	0.06

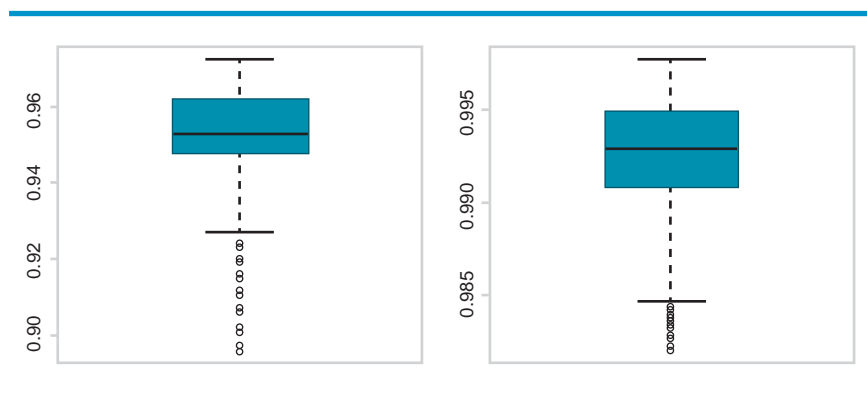
Fuente: elaboración propia.

Tabla 8. Eficiencia del estimador GREG ajustado.

f	10%			20%			50%		
P(S_A)	1%	50%	99%	1%	50%	99%	1%	50%	99%
Cor -0.9	46194.90	7908.93	6.32	62732.82	19962.41	18.58	40859.14	27452.96	13.03
Cor -0.7	46447.53	19184.00	3.95	62635.15	13471.40	7.49	41778.57	29787.39	31.28
Cor -0.5	59066.28	16481.67	8.99	8048.93	35692.92	6.53	65840.57	20758.00	4.53
Cor -0.3	27014.81	5768.08	2.17	23961.83	33193.49	4.95	40072.26	16423.36	8.08
Cor -0.1	7043.39	3973.29	7.61	9736.54	4420.25	6.40	12252.18	3746.25	26.23
Cor 0.1	5895.80	4825.34	5.15	5958.52	7206.76	4.50	17748.37	40262.71	19.32
Cor 0.3	27066.82	10797.53	3.76	20185.55	11179.54	4.45	18321.91	11343.88	13.81
Cor 0.5	48066.42	9140.47	5.70	21291.74	27110.42	6.02	48244.46	23276.30	9.89
Cor 0.7	37625.47	10243.33	5.57	64966.81	19681.22	6.46	42810.06	23083.71	28.40
Cor 0.9	44339.74	9338.89	16.51	49210.34	13601.37	4.68	44905.35	40940.08	8.43

Fuente: elaboración propia.

Gráfico 1. Diagrama de cajas para las probabilidades estimadas (izquierda) $\hat{\varphi}_k$ y $\hat{\rho}_k$ (derecha).



Fuente: elaboración propia.

- Edad (EDA): es la variable auxiliar Z_1 para realizar las estimaciones de calibración y regresión, que es el número de años cumplidos.
- Ingreso total (INGR): es la característica de interés Y , la cual es el ingreso total recibido por la persona.

Además de lo anterior, se tiene en cuenta que el tamaño de muestra en la base de datos GEIH es de 15266. En particular, para la variable INGR, se evidencia que existe ausencia de respuesta en 714 elementos. Para este escenario se aplica el método de *propensity score* mediante la regresión logística, quedando definida la estimación de la probabilidad de respuesta de la siguiente manera:

$$\hat{\varphi}_k = \frac{e^{\hat{\beta}_1 X_{1k} + \hat{\beta}_2 X_{2k} + \hat{\beta}_3 X_{3k}}}{1 + e^{\hat{\beta}_1 X_{1k} + \hat{\beta}_2 X_{2k} + \hat{\beta}_3 X_{3k}}}$$

Donde los $\hat{\beta}$ son los coeficientes de regresión estimados. Para este caso particular, el modelo estimado para la probabilidad de respuesta fue:

$$\hat{\varphi}_k = \frac{e^{3.614X_{1k} + 3.5597X_{2k} + 0.055X_{3k}}}{1 + e^{3.614X_{1k} + 3.5597X_{2k} + 0.055X_{3k}}}$$

Siendo $\hat{\beta}$ todos los significativos. Con base en lo anterior, y una vez definidas todas las probabilidades de respuesta, se tiene que la estimación del ingreso total de la PEA con el estimador HT ajustado fue de 2,153043e+13, con la estimación de calibración fue de 2,153038e+13, y por último la estimación GREG ajustada fue de 2,152999e+13. Todos los anteriores valores son muy cercanos entre sí. En el gráfico 1 se muestran las probabilidades $\hat{\varphi}_k$ y las probabilidades $\hat{\rho}_k$. En el primer caso, se observa la existencia de ponderadores extremos por debajo de aproximadamente 0,925, el 75% se encuentra entre 0,925 y 0,980, y la mediana gira alrededor de 0,955. En el segundo caso, se observa la existencia de ponderadores extremos por debajo de aproximadamente 0,985, el 75% se encuentra entre 0,985 y 1, y la mediana gira alrededor de 0,993.

Por otro lado, para estimar la varianza de los estimadores del total de ingresos, se realizó, mediante un análisis de Bootstrap, la aproximación de la varianza de estos estimadores generando una estimación

empírica de la distribución muestral de cada estimador. Dado lo anterior, se tiene que la varianza del ingreso total de la PEA con HT ajustado fue de $8,520057e+21$, la estimación de calibración fue de $5,198439e+21$, y por último la estimación GREG ajustada fue de $4,998861e+21$. Con lo anterior, el coeficiente de variación estimado, definido como la raíz de la varianza estimada sobre la realización del estimador, para el HT ajustado fue de $4,287149e-03$, para el estimador de calibración ajustado fue de $3,348766e-03$, y por último para la estimación GREG ajustada fue de $3,283913e-03$.

6. Discusión

Está demostrado que la ausencia de respuesta puede tener grandes efectos en los resultados de una encuesta. Estrategias inmediatas como incrementar el tamaño de la muestra sin enfrentar la falta de respuesta no reduce el sesgo y tal vez se desperdician recursos que hubiesen servi-

do para reducir o remediar esto. Por esa razón, una muestra mayor solo proporciona más observaciones de la clase de personas que contestarían la encuesta y aumentaría la brecha con los individuos que no contestaron el cuestionario.

El análisis de los registros completos tiene como hipótesis la idea de que quienes no responden son similares a quienes sí, y que las unidades con elementos faltantes son similares a quienes tienen las respuestas de todas las preguntas. Muchas evidencias indican que estas hipótesis son erróneas; se debe tratar de obtener, entonces, al menos, cierta información acerca de quienes no respondan, que puede emplearse para realizar ajustes por la ausencia de respuesta.

Los resultados obtenidos en este trabajo fueron simulados mediante la ponderación ajustada a la probabilidad de respuesta calculada mediante regresión logística. La metodología realizada consistió en

efectuar un modelo de regresión logística entre el estado de respuesta (0 = no responde, 1 = responde) y un conjunto de variables disponibles a nivel muestral (o poblacional) para respondientes y no respondientes. Los valores obtenidos de la regresión se toman como estimaciones de la probabilidad de respuesta. El uso del *propensity score* con la estimación de GREG y de calibración ha demostrado ser uno de los métodos más eficaces, ya que tiene un buen comportamiento en situaciones en las que otros métodos fracasan. Si el porcentaje de valores faltantes es muy bajo, los métodos de *propensity score* proporcionan resultados muy similares a los de los demás métodos de asignación.

Por último, se nota que los estimadores propuestos asumen que existe una relación entre la característica de interés y las variables auxiliares. Sin embargo, es posible definir otro tipo de estimadores que tengan en cuenta el supuesto de no linealidad.

Bibliografía

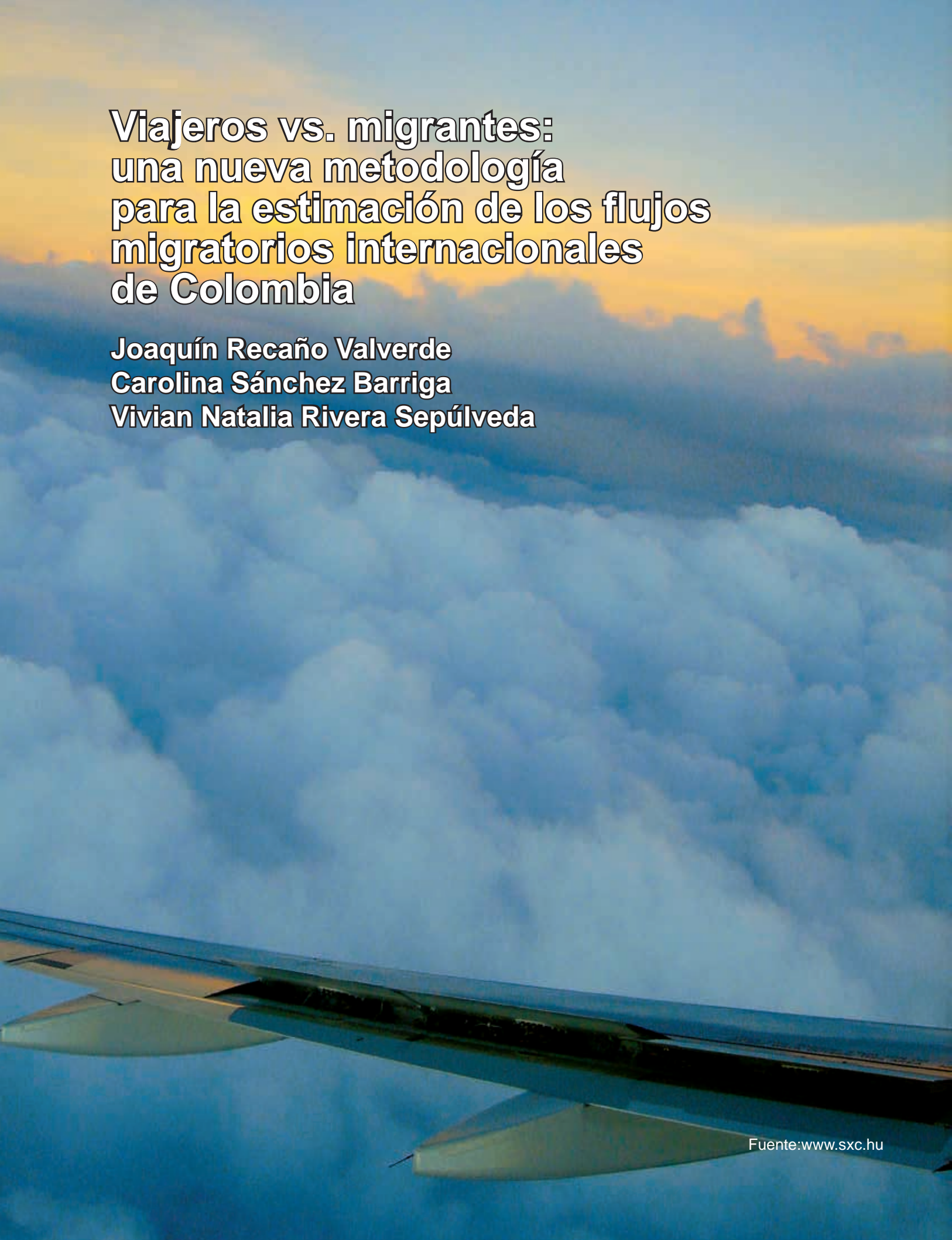
Guo, S. & Fraser, M. (2009). *Propensity Score Analysis*. SAGE Publications.

Gutiérrez, H. A. (2009). *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Universidad Santo Tomás.

Kim, J. K. & Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38, 157-165.

Lohr, S. L. (2010). *Sampling: design and analysis*. Cengage Learning.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.



Viajeros vs. migrantes: una nueva metodología para la estimación de los flujos migratorios internacionales de Colombia

**Joaquín Recaño Valverde
Carolina Sánchez Barriga
Vivian Natalia Rivera Sepúlveda**

Viajeros vs. migrantes: una nueva metodología para la estimación de los flujos migratorios internacionales de Colombia¹

Joaquín Recaño Valverde: Licenciado en Historia Contemporánea. Doctor en Geografía de la Universidad de Barcelona. DEA de Démographie et Sociétés de la Universidad de Paris-X, Francia. Profesor titular de Geografía Humana de la Universidad Autónoma de Barcelona e investigador asociado del Centro de Estudios Demográficos. Correo electrónico: jrecano@gmail.com

Carolina Sánchez Barriga: Estadística de la Universidad Nacional de Colombia. Profesional del Grupo de Migración de la Coordinación de Demografía de la Dirección de Censos y Demografía del DANE. Correo electrónico: csanchezb@dane.gov.co

Vivian Natalia Rivera Sepúlveda: Estadística de la Universidad Santo Tomás. Profesional en Estadística de la Universidad Antonio Nariño, Oficina de Autoevaluación y Acreditación. Correo electrónico: nataliaje33@gmail.com

Fecha de recepción: 31 de octubre de 2012
Fecha de aceptación: 10 de octubre de 2013

Resumen: en este artículo desarrollamos una metodología de estimación de los flujos migratorios internacionales de Colombia basados en una nueva fuente de información desarrollada por el Departamento Administrativo Nacional de Estadística (DANE) a partir de la base Migración Colombia y además presentamos los resultados obtenidos. Los datos de este estudio proceden de los flujos de 41.766.428 viajes internacionales registrados en el período 2004-2010. La metodología implementada permite reconstruir los itinerarios de viaje de 9.584.315 personas en el mismo período, quienes son sometidas a diferentes algoritmos para obtener su estatus migratorio. Finalmente, se presenta la metodología de estimación de los flujos migratorios y una comparación con los datos registrados en países como España, uno de los principales destinos de la emigración colombiana.

Palabras claves: migraciones internacionales, viajes internacionales, métodos de estimación, flujos migratorios Colombia.

Abstract: in this paper we present a methodology and the results of a new source of information on flows international migration Colombia developed by the National Administrative Department of Statistics (DANE) from the database Migración Colombia. The data for this study come from the

¹ Los datos estadísticos expuestos y analizados en el presente artículo fueron suministrados por la Unidad Administrativa Especial Migración Colombia, creada mediante Decreto Ley 4062 de 2011. Dicha entidad es la encargada de ejercer las funciones de autoridad y vigilancia y control migratorio y de Extranjería en el Estado colombiano.

flows of 41,766,428 international travels recorded during the period 2004-2010. The methodology used reconstructed the travel itineraries of 9,584,315 people in the same period, who are subjected to different algorithms to obtain the migration status. Finally, we present the methodology of estimation of migration flows and a comparison with the data recorded in countries like Spain, one of the main destinations of Colombian outmigration.

Keywords: international migration, international travel, estimation methods, migration flows Colombia.

1. Introducción

En este trabajo se presentan la metodología y los resultados de una estimación indirecta² de los flujos de migración internacional de Colombia. El proyecto, que está llevando a cabo el Departamento Administrativo Nacional de Estadística (DANE) a partir de la base de datos del registro de entradas y salidas internacionales administrado actualmente por la Unidad Administrativa Especial Migración Colombia, aprovecha los datos recogidos en 34 puestos de control migratorio a lo largo del país. El estudio se encuentra en la actualidad en una fase avanzada de trabajo; no obstante, los datos que aquí se muestran son preliminares y no constituyen los definitivos de la estadística que se persigue. Se ha considerado oportuno que la comunidad científica conozca el desarrollo de esta nueva fuente de información sobre migraciones internacionales, con la que se persigue solventar los importantes vacíos que al respecto existen sobre datos de migración internacional en Colombia.

Como es bien conocido, en un número muy importante de países no existe la posibilidad de contabilizar las entradas y salidas internacionales de un territorio. Colombia no dispone en la actualidad de un instrumento estadístico específico que mida directamente la migración exterior; sin embargo, la fuente administrativa desarrollada para el control de fronteras puede

convertirse mediante los procedimientos apropiados en la base de información para un nuevo registro que mida de forma indirecta la emigración e inmigración internacional.

Los aspectos que serán tratados en este artículo son los siguientes:

- a. En primer lugar se efectúan diversas consideraciones sobre el empleo del lugar de nacimiento como variable analítica y se establece la definición de migración.
- b. En segundo lugar se realiza una serie de consideraciones sobre la naturaleza de la información contenida en la base de viajeros de Migración Colombia y se presenta la definición de migrante/migración basada en un esquema espacio-temporal homologable con la definición de Naciones Unidas (1999).
- c. En tercer lugar se muestran las transformaciones necesarias en la base de datos desde la información inicial de viajes a otra de viajeros y la nueva distribución de frecuencias, y se analiza de nuevo la coherencia de la base de datos a través de las secuencias coherentes de viaje.
- d. En cuarto lugar, se presentan los fundamentos del algoritmo de estimación de los flujos migratorios.
- e. Finalmente, se analiza la consistencia del algoritmo desarrollado mediante la comparación de los resultados de la estimación de flujos emigratorios hacia España con la me-

dición de flujos de inmigrantes colombianos que realizan los microdatos de la Estadística de Variaciones Residenciales para el período 2004-2009 en sus dimensiones temporales y de sexo y edad. También se recoge una recopilación de flujos inmigratorios de una serie de países, que proporciona Naciones Unidas, que serán objeto de examen.

2. El contexto y el concepto de migración internacional

Antes de avanzar en la descripción de la metodología de estimación de flujos debemos definir claramente algunos conceptos que se van a emplear a lo largo del trabajo. El primero de ellos atañe a quién se considera un *colombiano en el exterior*. El segundo se refiere a cuál es el concepto de migración empleado en este trabajo y que está relacionado con las dificultades de establecer definiciones sobre los conceptos migratorios a partir de la naturaleza espacio-temporal del fenómeno (Courgeau, 1973; Naciones Unidas, 1999; Poulain, 1985; Willekens, 1985).

Algunos países se han centrado tradicionalmente en la producción de datos sobre los residentes extranjeros (países europeos), mientras que otros, los denominados países de nueva inmigración (Australia, Canadá, Nueva Zelanda y Estados Unidos), se han centrado en las estadísticas de nacidos en

² La base Migración Colombia permite una estimación indirecta de los flujos pero no de los stocks de población, ya que para una estimación de efectivos es necesario apoyarse en un recuento independiente.

el extranjero con independencia de su estatus de nacionalidad. Esta importante diferencia de enfoque se refiere a la historia de los sistemas de inmigración y a la legislación sobre ciudadanía y naturalización que han regido esos países. Para las sociedades receptoras, la población nacida en el extranjero es la representación de la primera generación de inmigrantes³.

El tamaño y la composición de la población nacida en el extranjero se determinan por la evolución de los flujos inmigratorios y la mortalidad; dicha población disminuye cuando las entradas de inmigrantes decrecen. En cuanto al concepto de población extranjera, se puede incluir a personas nacidas en el extranjero que mantuvieron la nacionalidad de su país de origen pero también a las generaciones nacidas en el país anfitrión que por limitaciones en el derecho de ciudadanía no adquieren la nacionalidad hasta pasado un período temporal determinado. Esta distinción da lugar a numerosas confusiones conceptuales a la hora de medir el impacto real de la emigración exterior en las sociedades emisoras.

Las características de la población de nacionalidad extranjera y de los flujos que las modifican dependerán, por lo tanto, de una serie de factores: el primero de ellos la historia de los flujos migratorios, en segundo lugar el crecimiento natural de la población denominada o percibida como «extranjera» y, finalmente, los procesos de naturalización. No obstante, como señalan la mayor parte de los especialistas en la materia, es preferible realizar un análisis basado en el lugar de nacimiento porque es una característica que permanece invariable a lo largo de todo

el período observado; de esta forma, las estadísticas elaboradas en distintos momentos del tiempo son plenamente comparables. Esta reflexión viene a cuenta porque la mayor parte de la información que proporcionan los países se refiere a recuentos censales o encuestas que nos muestran información sobre *stocks* de extranjeros. Sin embargo, construir una estadística sobre flujos migratorios significa establecer previamente el estatus migratorio de las personas para luego medir los flujos, lo que nos retrae al viejo problema del migrante-migración (Courgeau, 1973). Un migrante puede realizar varias migraciones; la naturaleza de estas últimas vendrá determinada por los cambios en el lugar residencial y la dimensión temporal del movimiento efectuado.

En este trabajo el análisis se centrará en la emigración exterior de Colombia. La razón de esta selección radica en que los sistemas de validación de nuestra estimación con la base Migración Colombia

se hallan en los países de destino. Para evaluar la pertinencia de la estimación utilizaremos las series que recopilan las Naciones Unidas de flujos de una serie de países. Las definiciones de estos flujos (por nacionalidad, país de nacimiento, país de residencia) varían de un país a otro, por lo que hemos tenido que adaptar nuestras tabulaciones a las determinadas por cada destino de la emigración que se compara⁴.

Abordemos ahora la definición de migración que emplearemos en este trabajo. Naciones Unidas en su manual de revisión de 1999 sobre las migraciones internacionales establece dos aspectos que caracterizan la definición de ese tipo de movimientos (tabla 1): el cambio de país de residencia habitual y el establecimiento durante al menos 12 meses fuera del país de residencia anterior (denominado como migrante por largo plazo); ambas condiciones son las que se han tenido en cuenta para estimar la migración internacional en la base de datos de Migración Colombia.

Tabla 1. Definición de Naciones Unidas sobre migración internacional.

Definiciones de país de residencia habitual, de migrante internacional por largo plazo y migrante internacional por breve plazo	
País de residencia habitual	El país en el que la persona vive, es decir, el país en el que tiene una vivienda donde normalmente pasa los períodos diarios de descanso. Los viajes temporales al exterior con fines de ocio, vacaciones, visitas a parientes y amigos, negocios, tratamiento médico o peregrinación religiosa no cambian el país de residencia habitual.
Migrante por largo plazo	Toda persona que se traslada, por un período de por lo menos un año (12 meses) a un país distinto de aquél en el que tiene su residencia habitual, de modo que el país de destino se convierte efectivamente en su nuevo país de residencia habitual. Desde la perspectiva del país de partida la persona será un emigrante por largo plazo y desde el punto de vista del país de llegada, la persona será un inmigrante por largo plazo.

Fuente: Tomado de Naciones Unidas (1999), p. 10.

Para el desarrollo de esta estimación con nuestros datos se han considerado otras recomendaciones de Naciones Unidas, en especial las que van de la 70 a la 74 y que hemos adaptado a nuestros datos:

70. Permanencia de un año en el país de salida u origen y forma de comprobarlo.

72. La existencia de una defini-

ción a priori del carácter de migrante internacional en la estadística de frontera; en este caso, Migración Colombia. El algoritmo de explotación sigue dichas recomendaciones.

73. El uso de un formulario de entrada y salida. En este caso es un formulario electrónico que se integra en una base de datos que se nutre desde todos los

³ Numerosos trabajos señalan esta distinción fundamental entre la primera y la segunda generación de inmigrantes, llegando a definir la generación 1.5 como aquella que afecta a las personas nacidas en el extranjero que llegaron antes de los 10 años al país de acogida.

⁴ Las características de cada una de las series de datos pueden verse en <http://esa.un.org/MigFlows/Data%20availability.html>

puntos de frontera de Colombia: aéreos, terrestres y marítimos.

74. El cotejo de base de datos permite enlazar entradas y salidas de un mismo individuo en diferentes momentos del tiempo; se trata de una reconstrucción longitudinal de los viajes y migraciones de la población objeto de análisis.

75. Se ha especificado claramente el tipo de criterio cronológico en la elaboración de la estadística: duración de estancia en el exterior o Colombia para cada individuo, superior a los 12 meses.

3. Metodología de estimación de los flujos de entradas y salidas internacionales de Colombia con la base Migración Colombia

3.1. La información proporcionada por la base Migración Colombia.

La base Migración Colombia es un registro administrativo compuesto por 29 campos de información cuyas variables más significativas se describen en la tabla 2. Estas informaciones son proporcionadas por los viajeros a través de las preguntas efectuadas en el control fronterizo y mediante la lectura mecánica del documento de entrada. Se trata de información demográfica, temporal y geográfica suficiente para la aplicación de una definición robusta de la migración, consistente con las demandas de la definición de Naciones Unidas.

El registro de Migración Colombia mide flujos de viajeros; solo una parte de esos viajes acabarán siendo definidos como migraciones. Por ejemplo, una persona identificada en esa base puede realizar

durante el período 2004-2010 innumerables viajes, pero difícilmente superará las dos o tres migraciones internacionales en los casos más extremos. Los factores que distinguen un concepto del otro radican en el intervalo temporal y en la definición de residencia. El evento denominado *viaje* es independiente del lugar de residencia del viajero; por el contrario, la definición de migración necesita estimar un lugar de residencia que se

modifica con el evento migratorio y un intervalo temporal determinado que no se tiene en cuenta en la definición del evento *viaje*. Las categorías analizadas, empleadas en este trabajo, distinguen entonces entre flujos/eventos y *stocks*/personas, siguiendo el esquema de la desigualdad de Courgeau (1973).

Tabla 2. Información recogida en la base Migración Colombia.

Tipo de variable	Descripción información Migración Colombia
Identificador individuo	Identificador individuo previamente anonimizado
Administrativas	Tipo de Documento
	Numero de Documento
	Numero de Viaje
	País expide pasaporte
	Punto de control
VISA	Medio de Transporte
	Tipo de Visa
	Fecha de expedición Visa
	País de Expedición Visa
	Numero de días máximos de permanencia
Tipo de flujo "viaje"	I entrada en Colombia, E salida de Colombia
Datación temporal	Fecha de viaje
Demográficas	Sexo
	Fecha de Nacimiento
	País de nacimiento
	Nacionalidad del Viajero
Geográficas	País de Escala
	País de Residencia
	País Destino
Socio-económicas	Ocupación
	Motivo viaje

Fuente: Migración Colombia. Elaboración propia.

Uno de los principales problemas que se presentan cuando se utilizan datos de registros de viajeros radica en la confusión conceptual entre viaje y migración y entre migrante y viajero (tabla 3). Esto lleva a pensar que la diferencia entre los viajes de entrada y los viajes de salida de Colombia son realmente un *proxy* de los saldos migratorios. Nada más lejos de la realidad, ya que en los viajes no existe ninguna referencia temporal de estadía en

Colombia o fuera de ella, que es la variable que determina el estatus migratorio de las personas. Todo sistema de registro temporal recoge, desde la perspectiva de la demografía, flujos o eventos.

En el siguiente apartado se describen los procesos a los que hemos sometido la base Migración Colombia para convertir la información de viajes/viajeros en migraciones/migrantes.

Tabla 3. Distinción evento/*stock* en viajes/migración.

Eventos		Stocks
Viajes	≥	Viajeros
Flujos	≥	Personas
Migraciones	≥	Migrantes
Registro	≥	Censos
Viajes	≥	Migraciones

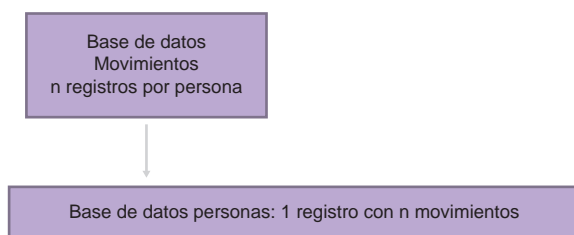
Fuente: Elaboración propia siguiendo las indicaciones de Courgeau (1973).

3.2. Tratamiento informático realizado en la base Migración Colombia.

Los datos de viajeros registrados en control de fronteras son sometidos a un proceso de anonimización que asigna un código identificador unívoco para cada individuo presente en la base de datos Migración Colombia. A partir de aquí se

procede a la constitución de una nueva base de viajeros siguiendo un modelo de transposición (diagrama 1). Esto supone pasar de un fichero con *n registros para cada individuo* (en términos de base de datos, un registro para cada viaje con un identificador de individuo) a un fichero con un *individuo con n campos correspondientes a las informaciones de cada viaje*.

Diagrama 1. Transposición de la base de viajes de Migración Colombia en una base de viajeros.



Fuente: elaboración propia.

Una vez constituidas las secuencias de viajes de cada individuo (viajero), reconstruidas a partir del código unívoco de identificación, se procede al análisis de la coherencia de la secuenciación de viajes, de tal manera que a una salida le sigue una entrada y viceversa. Veamos diferentes ejemplos de secuenciación:

Un solo episodio para un viajero en la base MC

Tipo 1: Un solo episodio de salida (E). Con las limitaciones de observación temporal señaladas anteriormente.

Tipo 2: Un solo episodio de entrada (I). Idéntico planteamiento.

Un viajero con dos viajes del tipo E-I

Tipo 3a (E-I): Si el intervalo temporal entre E e I es inferior a un año, se tratará de un viajero.

Tipo 3b: E-I (mayor de 12 meses). Se trata de un emigrante según la definición Naciones Unidas, puesto que el intervalo temporal entre E e I es superior a un año.

La misma consideración se aplica a las secuencias I-E.

Episodios más complejos

E-I-E: Corresponde a una salida, retorno y posterior salida. Los in-

tervalos temporales definirán su carácter migratorio.

I-E-I-E: Puede darse en la base de datos la observación truncada, ya que esta comienza en 2004; la primera entrada puede obedecer a una emigración previa, anterior a 2004, no controlada en la base de datos actual.

Episodios incompletos o incorrectos

E-I-E-E

I-I

Ambos episodios no son coherentes. En el primero de los casos podemos deducir que existe un cuarto viaje (I) no identificado. En el otro caso existe una salida de Colombia que se efectuaría en segundo lugar. Estos episodios de secuenciación incorrecta no son contemplados en nuestro análisis actual, pero serán objeto de análisis específicos en próximos trabajos.⁵

Debemos subrayar que la aplicación de este tipo de análisis a la coherencia de las secuencias de viajes es inédita. A diferencia del análisis estadístico tradicional que persigue la depuración de los viajes como variables independientes sin tener en cuenta una secuenciación lógica de los itinerarios. Por el contrario, en este trabajo se va a considerar siempre el conjunto del itinerario de viajes de una misma persona y su adecuación o coherencia. El análisis de la secuenciación lógica de los itinerarios de viajeros tiene como objetivo estimar los potenciales errores en el registro de entradas y salidas desde una perspectiva demográfica y migratoria. Por otra parte, se trata de verificar si los puntos de entrada o salida terrestre son el origen de alguna de las secuencias truncadas del tipo E-E o I-I. Se puede constatar en la tabla 4 que parte de los problemas de fiabilidad del registro (itinerarios truncados o incorrectos) se encuentran localiza-

⁵ El número de secuencias incompletas o incorrectas y personas se sitúa en torno al ocho por ciento.

dos en aquellos individuos que pasan en algún momento por puntos de control de naturaleza marítima y/o terrestre, en especial cuando los individuos presentan itinerarios con medios de transporte mixto (en los modelos marítimos- terrestres uno

de cada dos viajeros muestran itinerarios incoherentes [49,9%]). Por el contrario, las secuencias de itinerarios únicamente aéreos presentan una elevada proporción de secuencias coherentes de casi el 94%.

La base B, en la que participa el 1,65% del conjunto de viajeros supone el 21,9% de todos los viajes⁷.

En lo que se refiere al fichero A, que es el que interesa a efectos analíticos, este contenía 28.112 secuencias de viajes de las cuales, 50 eran coherentes y 28.062 incoherentes. De las 50 combinaciones lógicas posibles, 25 corresponden a itinerarios iniciados con una entrada a Colombia (I: 25) y los otros 25, a episodios que comienzan con una salida de Colombia (E: 25). No obstante, existe un número mucho más elevado de secuencias inconsistentes o truncadas, en las que a una E (salida) o I (entrada) no le sucede una I (entrada) o E (salida)⁸. La distribución de secuencias lógicas se puede observar en la tabla 5. Un indicador que refuerza la consistencia y fiabilidad de la BD es que un 91,7% de las personas del fichero A tienen itinerarios coherentes y solo un 9,3% poseen secuencias no coherentes (muchas de ellas, como hemos subrayado, por la incorporación de itinerarios no aéreos o mixtos).

Tabla 4. Itinerarios lógicos y secuencias individuales de transporte. 2004-2010.

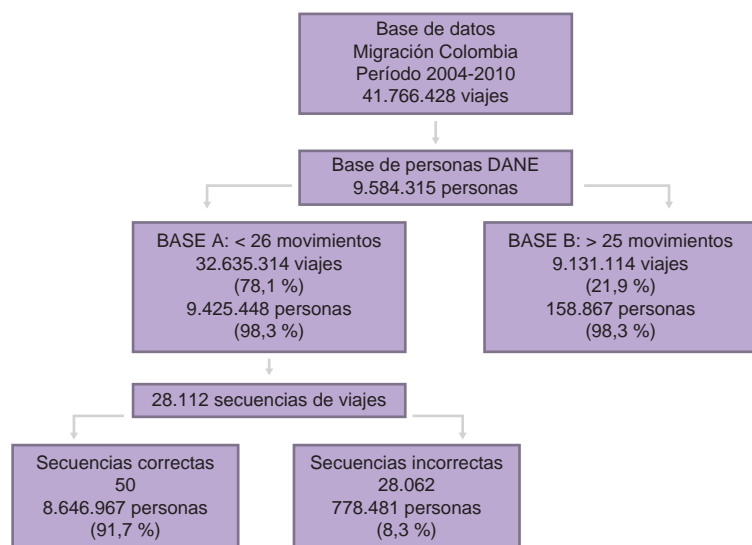
Medio de transporte utilizado	Viajeros	% Secuencias	% Secuencias coherentes
Aéreo	7.824.906	83,02%	93,9%
Terrestre	1.252.564	13,29%	83,8%
Marítimo	255.457	2,71%	65,4%
Mixtos			
Aéreos-Terrestres	41.598	0,44%	97,2%
Aéreos-Marítimos	46.097	0,49%	81,7%
Marítimos-Terrestres	1.880	0,02%	51,1%
Aéreos-Marítimos-Terrestres	2.946	0,03%	54,8%
Total	9.425.448	100,00%	91,7%

Fuente: Elaboración propia a partir de la base Migración Colombia⁶.

Tras realizar la conversión de viajes en viajeros, se encontró que un número no desdeñable de personas habían efectuado más de 25 viajes en la base de Migración Colombia a lo largo del período 2004-2010; incluso algunas superaban los 500 viajes. Se trata de viajeros frecuentes y en su mayor parte tripulantes de aeronaves que no eran suficientemente identificados en esa base. Este constituye en la actualidad uno de los problemas del registro. Para abordar de una forma eficiente el análisis de esta información se procedió a la división de la BD de personas en dos ficheros, tal como se describe en el diagrama 2.

personas que habían efectuado menos de 26 movimientos en el período 2004-2010 (diagrama 2), las cuales representan en la base de Migración Colombia el 98,3% de las personas y el 78,1% de todos los movimientos en ambas bases.

Diagrama 2. Distribución de secuencias en la base de personas que elabora el DANE.



⁶ En esta tabulación no se consideran las personas que realizaron más de 25 viajes a lo largo del período 2004-2010, por tratarse en buena parte de los casos de integrantes de tripulaciones aéreas o itinerarios terrestres.

⁷ Los ficheros de tipo B serán objeto de un análisis particular del que hay que extraer la información correspondiente a las tripulaciones, procedimiento aún no depurado por los nuevos responsables de la base Migración Colombia que sustituye al DAS.

⁸ En total existen 28.062 secuencias que no serían incorrectas porque en algunas oportunidades, en zona fronteriza, los ciudadanos o autoridades obvian el control migratorio, lo que puede generar un salto en la secuenciación prevista de entradas y salidas.

Fuente: Elaboración propia a partir de la base Migración Colombia.

Un total de 3.533.113 personas disponían de secuencias coherentes iniciadas con una salida de Colombia (36,3% de las personas), correspondientes a 12.168.654 viajes (29,13% de la base de viajes). Un 63,8% de estas secuencias lo constituyen secuencias de episodios de rango 1 y 2 (equivalentes al número de viajes), y el acumulado hasta el rango 5 suponía el 82,3% de las personas que

inician sus itinerarios coherentes con una salida (gráfico 1).

Las personas que inician con una secuencia lógica del tipo entrada en Colombia (I) fueron 5.113.854. En la base las secuencias de viajes de rango 1 y 2 representan el 42% y el acumulado hasta el rango 5 supone el 68,8% de las personas que inician sus itinerarios coherentes con una entrada. Los itinerarios

con inicio de entrada en Colombia corresponden, en buena parte, a visitantes turísticos y de negocios (el viaje del tipo I-E constituye el 42% de las personas y casi el 40% de los viajes en este tipo de secuencias) (figura 4).

En el próximo apartado aplicaremos la definición de migración de Naciones Unidas a los datos de la base Migración Colombia.

Tabla 5. Secuencias viajeras iniciadas con una salida o entrada desde Colombia. 2004-2010.

Rango de secuencia	Tipo Secuencia	Personas	Viajes	% personas	Tipo Secuencia	Personas	Viajes	% personas
1	E	1.061.021	1.061.021	30,00%	I	853.494	853.494	7,30%
2	EI	1.192.233	2.384.466	33,70%	IE	2.927.115	5.854.230	42,00%
3	EIE	192.126	576.378	5,40%	IEI	134.011	402.033	2,80%
4	EIEI	373.871	1.495.484	10,60%	IEIE	562.117	2.248.468	14,70%
5	EIEIE	88.337	441.685	2,50%	IEIEI	60.908	304.540	2,00%
6	EIEIEI	180.750	1.084.500	5,10%	IEIEIE	212.610	1.275.660	7,80%
7	EIEIEIE	48.866	342.062	1,40%	IEIEIEI	35.180	246.260	1,60%
8	EIEIEIEI	104.390	835.120	3,00%	IEIEIEIE	104.312	834.496	4,90%
9	EIEIEIEIE	29.789	268.101	0,80%	IEIEIEIEI	22.096	198.864	1,20%
10	EIEIEIEIEI	65.254	652.540	1,80%	IEIEIEIEIE	57.208	572.080	3,30%
11	EIEIEIEIEIE	19.125	210.375	0,50%	IEIEIEIEIEI	14.756	162.316	1,00%
12	EIEIEIEIEIEI	43.139	517.668	1,20%	IEIEIEIEIEIE	34.378	412.536	2,30%
13	EIEIEIEIEIEIE	13.069	169.897	0,40%	IEIEIEIEIEIEI	10.113	131.469	0,80%
14	EIEIEIEIEIEIEI	29.774	416.836	0,80%	IEIEIEIEIEIEIE	22.266	311.724	1,70%
15	EIEIEIEIEIEIEIE	8.876	133.140	0,30%	IEIEIEIEIEIEIEI	7.123	106.845	0,60%
16	EIEIEIEIEIEIEIEI	20.934	334.944	0,60%	IEIEIEIEIEIEIEIE	14.401	230.416	1,30%
17	EIEIEIEIEIEIEIEIE	6.270	106.590	0,20%	IEIEIEIEIEIEIEIEI	4.988	84.796	0,50%
18	EIEIEIEIEIEIEIEIEI	15.249	274.482	0,40%	IEIEIEIEIEIEIEIEIE	10.149	182.682	1,00%
19	EIEIEIEIEIEIEIEIEIE	4.647	88.293	0,10%	IEIEIEIEIEIEIEIEIEI	3.682	69.958	0,40%
20	EIEIEIEIEIEIEIEIEIEI	11.506	230.120	0,30%	IEIEIEIEIEIEIEIEIEIE	7.115	142.300	0,80%
21	EIEIEIEIEIEIEIEIEIEIE	3.383	71.043	0,10%	IEIEIEIEIEIEIEIEIEIEI	2.790	58.590	0,30%
22	EIEIEIEIEIEIEIEIEIEIEI	8.855	194.810	0,30%	IEIEIEIEIEIEIEIEIEIEIE	5.328	117.216	0,60%
23	EIEIEIEIEIEIEIEIEIEIEIE	2.566	59.018	0,10%	IEIEIEIEIEIEIEIEIEIEIEI	2.151	49.473	0,30%
24	EIEIEIEIEIEIEIEIEIEIEIEI	6.994	167.856	0,20%	IEIEIEIEIEIEIEIEIEIEIEIE	3.920	94.080	0,50%
25	EIEIEIEIEIEIEIEIEIEIEIEIE	2.089	52.225	0,10%	IEIEIEIEIEIEIEIEIEIEIEIEI	1.643	41.075	0,20%
Total		3.533.113	12.168.654	100,00%	Total	5.113.854	14.985.601	100,00%

Fuente: Elaboración propia a partir de la base Migración Colombia.

3.3. Definición de la migración internacional en la base Migración Colombia.

La definición de migrante pivota sobre la aplicación de dos ópticas simultáneas, una *geográfica* derivada del lugar de residencia y otra *tempo-*

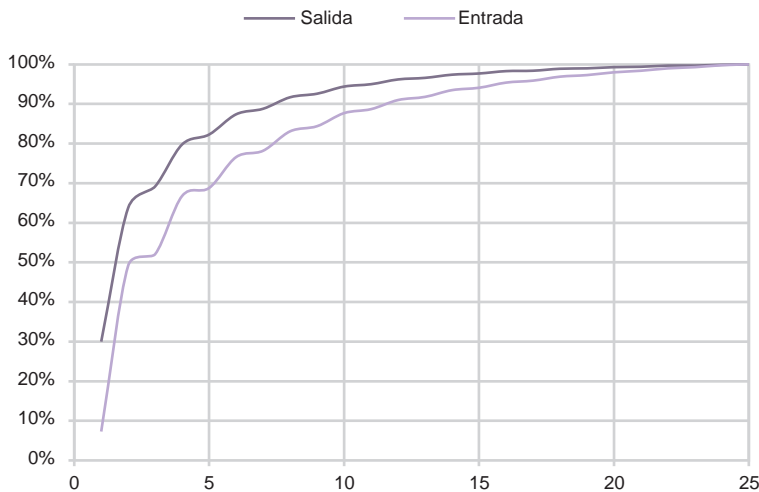
ral que miden la duración de estancia en Colombia o fuera de Colombia entre dos observaciones de viajes en los momentos t_1 y t_2 . De esta forma, en la base Migración Colombia hemos definido como migrante internacional a toda persona (viajero) que

realice un viaje internacional desde o hacia Colombia y permanezca más de un año fuera de Colombia o en Colombia. La duración de estancia se estima a partir de la comparación entre dos fechas de viajes sucesivas declaradas en ese registro⁹. Pero un

migrante puede realizar a lo largo del período observado más de una migración; esta óptica nos exige in-

troducir el concepto del flujo emigratorio o inmigratorio.

Gráfico 1. Secuencias lógicas acumuladas iniciadas por salida (E) o entrada (I) de Colombia. 2004-2010.



Fuente: Elaboración propia a partir de la base Migración Colombia.

Un ejemplo de una *emigración* en la base Migración Colombia es el siguiente: Si la persona tiene un movimiento de salida (viaje), identificado en la base de datos como «E», y este se realiza el 01/01/2004, solo se podrá considerar una emigración bajo dos supuestos (independientes):

- El individuo no regresa (no se identifica su entrada en Colombia en todo el período analizado).
- El individuo regresa a Colombia después del 01/01/2005.

En ambos casos debe transcurrir un período de un año para que el viaje se considere una emigración, no obstante la fecha del evento emigración será el 01/01/2004. Existen problemas adicionales asociados a esta definición. Puesto que la observación de la base Migración

Colombia está truncada temporalmente a la izquierda (solo observamos a los individuos a partir del 1 de enero de 2004) no conocemos cuál es la residencia del individuo a 1 de enero de 2004 (en Colombia o fuera de Colombia), al menos que lo declare correctamente en el registro¹⁰. Una salida E en esa fecha puede ser precedida de una entrada en el mes de diciembre de 2003 que no habrá sido observada en la actual base de datos, por lo que no podemos empezar a considerar las emigraciones en la base de datos hasta que transcurra un año fuera del país, a partir del 1 de enero de 2005¹¹. En el caso precedente el individuo habría realizado una secuencia del tipo I-E. Similar problema se produce en el truncamiento a la derecha: la observación deberá finalizar a 31 de diciembre de 2009 para los datos del período

2004-2010. El país de destino de la emigración se deriva de la información proporcionada por el viajero sobre el destino del viaje.

Por otra parte, en la base Migración Colombia se considera una *inmigración* el siguiente ejemplo: Si la persona tiene un movimiento de entrada, identificado en la base de datos como «I» y este se realiza el 02/08/2005, solo se podrá considerar una inmigración si el individuo no vuelve a salir de Colombia (no se identifica su salida de Colombia) o el individuo sale de Colombia después del 02/08/2006. En ambos casos debe transcurrir un período de un año para que el viaje se considere una inmigración, aunque la fecha de inmigración será el 02/08/2005. De forma similar a lo que ocurre en la emigración, el año 2004 y el 2010 en la base de datos tienen truncamientos de observación que impiden la consideración de todos los movimientos realizados en esos años en el análisis. El país de origen de la inmigración se deriva de la información proporcionada por el viajero sobre el origen del viaje.

4. Validación de la estimación de los flujos de emigración internacional realizados a partir de la base Migración Colombia

A lo largo de este trabajo, hemos ensayado la transformación de una base de información sobre viajes y

⁹ Existen algunos problemas en la identificación del país de residencia y su empleo directo para la estimación de la migración. Por ejemplo, muchos colombianos que llevan fuera de Colombia según la base de datos más de 12 meses siguen declarando como país de residencia Colombia cuando ingresan en el país en un viaje de corta duración. Por otra parte, los viajeros frecuentes cuyo período entre dos estancias en Colombia no supera los 12 meses aunque residan ya fuera de Colombia son de difícil captación en este primer análisis. Este grupo específico será objeto de análisis en próximos desarrollos de esta investigación. Por ejemplo, en el conjunto de la base sólo un 9.7 por ciento declara haber cambiado de país de residencia entre la información que considera los 25 primeros viajes. Es por ello que en la definición de migración sólo consideramos dos posibles residencias en Colombia y fuera de Colombia y ambas se desprenden directamente de la duración de estancia entre dos viajes sucesivos.

¹⁰ Este es uno de los principales retos con el que nos enfrentamos en la actualidad: establecer la fiabilidad del estatus migratorio del individuo al iniciar la observación en la base.

¹¹ La base de Migración Colombia no se ocupa de las emigraciones realizadas con anterioridad a 2004, ya que su sistema de imputación registra los flujos con duraciones superiores a un año que están presentes en la propia base de datos. Así, una persona que hubiera emigrado de Colombia en 2001 y entrara en Colombia en 2004 por vez primera y permaneciera más de un año sería considerada a todos los efectos como un flujo inmigratorio en 2004. A partir de ese momento cualquier movimiento de salida con permanencia superior a un año sería considerado ya como una emigración.

viajeros en una estadística aproximativa de los movimientos migratorios internacionales de Colombia; pero ¿son válidos estos resultados? Para efectuar la validación de las estimaciones realizadas hemos aplicado un test; este consiste en la comparación de los flujos emigratorios de la base Migración Colombia con los flujos de inmigración de colombianos medidos en los principales países de destinos de la emigración colombiana de los que se dispone de alguna información a partir de las fuentes que estos confeccionan. De esta forma, las salidas que hemos estimado a partir de la BD Migración Colombia deberían corresponder aproximadamente a los valores obtenidos en la medición en destino. El test se ha aplicado en varias fases con especial énfasis en los datos españoles para los cuales se dispone del acceso a los microdatos.

4.1. Comparación con los datos españoles

En la primera fase, los datos que se muestran en la tabla 6 realizan una comparación con cuatro explotaciones del registro de inmigración internacional derivado de la Estadística de Variaciones Residenciales (EVR) de España, que definimos a continuación:

a. La estadística de las inmigraciones internacionales de extranjeros, cuyo país de procedencia es Colombia, padece un cierto subregistro en España, ya que una parte importante de los extranjeros no declaran el país de procedencia cuando se empadronan; este resultado supone a efectos comparativos que las emigraciones estimadas por el método aplicado deberán ser –como así suce-

de– las más próximas a las registradas en España por país de procedencia. El método de estimación de flujos que hemos aplicado registra un 92,2% de las entradas de este tipo que muestra la EVR española.

b. Las inmigraciones que recibe España con Colombia como país de nacimiento son una estadística que registra cierta inflación en España para el propósito que perseguimos. Dos son las causas:

i) En primer lugar, algunos individuos nacidos en Colombia pueden llegar a España desde países terceros (no son registrados, por lo tanto, en la base Migración Colombia como migrantes con destino España sino como otro destino del espacio Schengen)¹²;

Tabla 6. Estimación de la emigración de Colombia con destino a España en el período 2004-2009 comparada con los registros españoles de la EVR del mismo período.

Flujos estimados con Migración Colombia	2004	2005	2006	2007	2008	2009	Total
País de procedencia (Colombia)	24.479	16.850	23.908	30.709	27.362	18.042	141.350
País de nacimiento	24.272	16.556	23.462	30.168	26.474	16.793	137.725
País de nacionalidad	24.083	16.506	23.467	30.125	26.308	16.757	137.246
País de nacionalidad y nacimiento	24.016	16.414	23.314	29.963	26.180	16.583	136.470
Todas las nacionalidades	27.290	18.031	25.147	32.846	30.231	21.680	155.225
Flujos inmigratorios registrados en España (INE)	2004	2005	2006	2007	2008	2009	Total
País de procedencia (extranjeros)	16.610	20.541	27.864	35.690	33.873	18.749	153.327
País de nacimiento	22.836	26.457	37.095	43.156	43.248	26.483	199.275
País de nacionalidad	21.502	24.945	35.621	41.725	42.166	25.558	191.517
País de nacionalidad y nacimiento	22.969	26.604	37.262	43.364	43.516	26.718	200.433
% Estimación (a)/Datos INE España (cobertura)	2004	2005	2006	2007	2008	2009	Total
País de procedencia – Colombia (extranjeros)	147,38%	82,03%	85,80%	86,04%	80,78%	96,23%	92,19%
País de nacimiento	106,29%	62,58%	63,25%	69,90%	61,21%	63,41%	69,11%
País de nacionalidad	112,00%	66,17%	65,88%	72,20%	62,39%	65,56%	71,66%
País de nacionalidad y nacimiento	104,56%	61,70%	62,57%	69,10%	60,16%	62,07%	68,09%
Todos los movimientos	118,81%	67,78%	67,49%	75,74%	69,47%	81,14%	77,44%

Fuente: Elaboración DANE a partir del Registro de entradas y salidas de Migración Colombia de octubre de 2011 y los microdatos de la Estadística de Variaciones Residenciales del INE, 2004-2009.

ii) En segundo lugar, una práctica habitual de algunos viajeros colombianos (con duración

de estancia fuera de Colombia inferior a los 12 meses y, por lo tanto, no computados como emigrantes) y con un viaje inferior a los tres meses que permi-

te la visa turista de Schengen pueden registrarse en el Padrón de Población en España, lo que supone una entrada inmigratoria en España pero ninguna

¹² Así parece apuntar los resultados que muestran las tablas 6 y 7.

contabilización como emigrante en Colombia¹³. Esta medida registra en este caso un 69,1% de las inmigraciones que de este tipo contabiliza España.

c. Las inmigraciones que recibe España con Colombia como país de nacionalidad. Este tipo de estadística muestra un cierto subregistro a efectos comparativos, en especial de aquellos colombianos de nacimiento que hayan obtenido la nacionalidad española u otra y esta sea la de entrada en España. El registro

de Migración Colombia muestra en este caso un nivel de cobertura del 71,7%.

d. La suma de las estadísticas b y c, país de nacionalidad y/o nacimiento = Colombia. Estos datos suponen la máxima estimación de inmigración, ya que aúnan las dos ópticas de estudio de los colombianos: país de nacimiento y/o nacionalidad. El registro colombiano contabilizaría un 68,1% de los registrados con datos españoles.

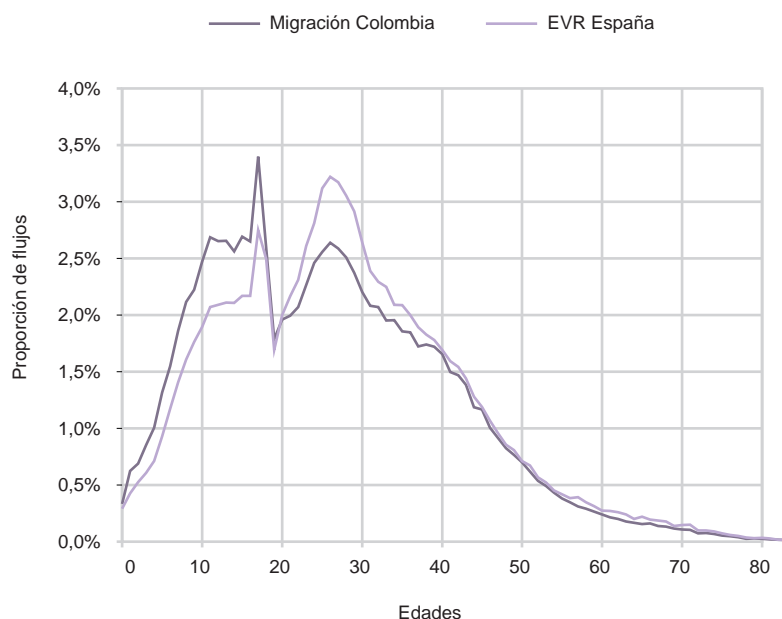
vo de los flujos emigratorios de la Base Migración Colombia con los correspondientes a 18 países que recoge la *International Migration Flows to and from Selected Countries: The 2010 Revision (web-based database)* de Naciones Unidas.

En su conjunto, la base Migración Colombia cubriría el 116.1% de los flujos que registran los países de destino. Teniendo en cuenta que la estimación de Migración Colombia puede mejorar aún, podemos decir que el nivel de cobertura es francamente satisfactorio. No obstante, este valor general oculta situaciones antagónicas. Así, en países de la Europa Central como Alemania, el nivel de cobertura es relativamente bajo, con un valor que apenas supera el 80%. La explicación reside probablemente en lo que sucede en España, Francia e Italia con niveles de cobertura superiores al cien por cien. Los países del sur de Europa serían los de destino primario, las puertas de entrada en Europa, que es el registrado en la información de la BD de Migración Colombia, mientras que otros países como los señalados en primera instancia serían destinos secundarios previo paso por el sur de Europa, en algunos casos con una regularización del estatus migratorio.

Pero, ¿qué pasa, por el contrario, con el sobre-registro de los países del Norte de Europa? Aquí la clave de la explicación radica en las estructuras por edades¹⁴. Si, como parece, el modelo de países nórdicos importa adopciones de colombianos, es posible que el tratamiento estadístico en los países receptores sea el de nacimiento y no de inmigración, que es el que detecta la base Migración Colombia.

Un caso diferente de elevado subregistro es el que presentan países como Canadá y Australia. Aquí es posible que la explicación tenga que ver con las duraciones entre dos via-

Gráfico 2. Estructura demográfica de las emigraciones colombianas hacia España a partir de la base Migración Colombia y flujos registrados en España por la Estadística de Variaciones Residenciales. 2004-2009.



Fuente: Elaboración propia a partir del Registro de entradas y salidas Migración Colombia y la Estadística de Variaciones Residenciales de España 2004-2009

e. Por otra parte, la estimación por edad (gráfico 2) es muy aproximada entre ambas

fuentes y es, por lo tanto, una certificación de la solidez de la estimación demográfica de los flujos que hemos implementado.

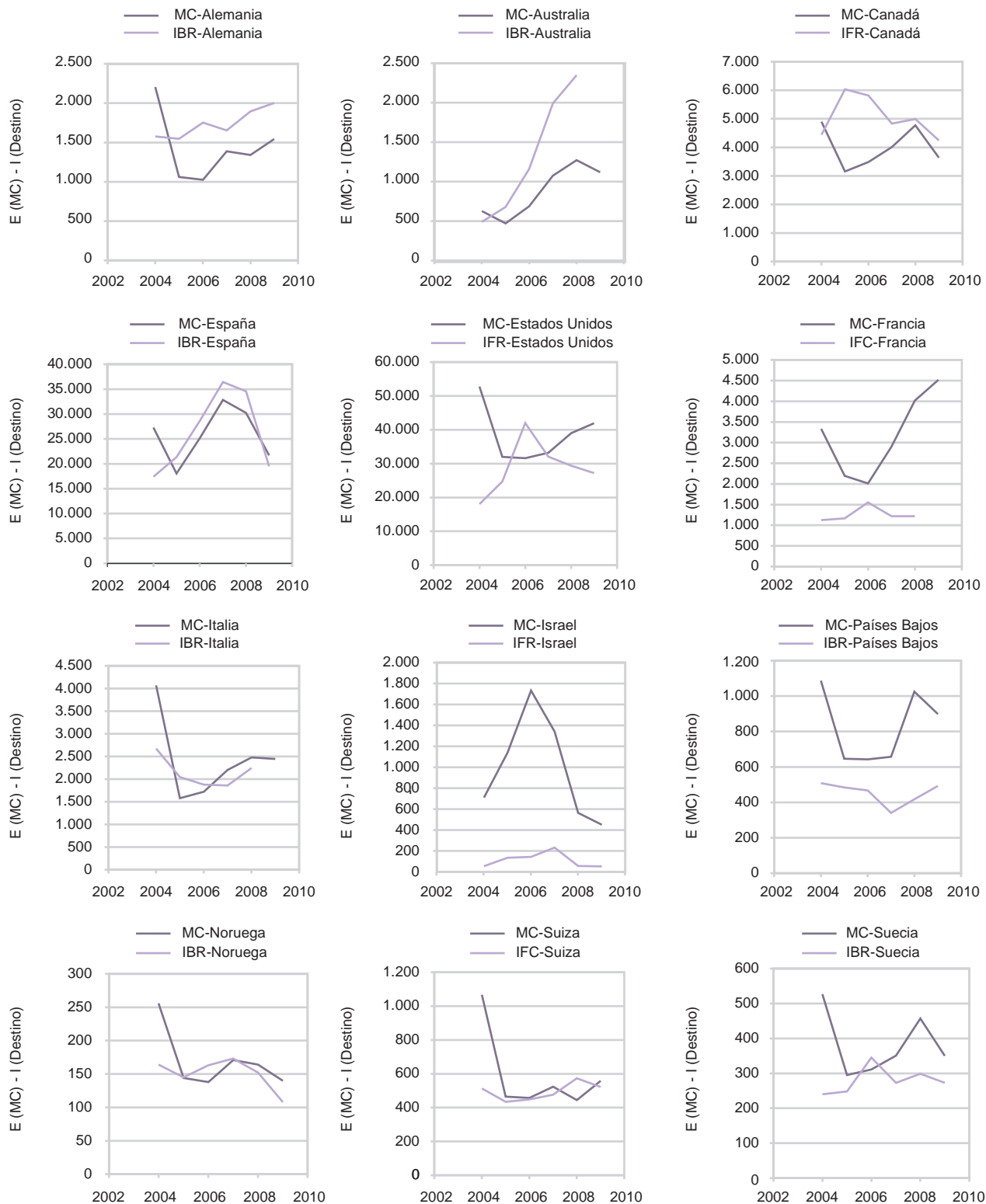
¹³ En el período 2006-2010, 42.051 individuos de nacionalidad colombiana fueron dados de baja por caducidad del registro del padrón; una parte importante de estos efectivos correspondían a individuos que habían salido de España sin notificarlo a las autoridades españolas y probablemente en situación administrativa irregular, lo que les otorga una más que probable estancia en España.

¹⁴ Ver documento de trabajo Una nueva base de datos para la estimación de los flujos migratorios internacionales de Colombia: Metodología y resultados comparativos (2012), anexos gráficos. Disponible en: http://www.dane.gov.co/candane/index.php?option=com_content&task=section&id=18 y anexo gráfico 1.

4.2. Comparación con los datos de flujos recopilados por Naciones Unidas.

En la tabla 7, anexo 1 de tablas y gráfico 3 realizamos un comparati-

Gráfico 3. Comparación de flujos emigratorios de la base Migración Colombia con los datos recopilados en los países de destino. 2004-2009.



Fuente: Elaboración propia a partir de la base Migración Colombia y United Nations, *International Migration Flows to and from Selected Countries: The 2010 Revision (web-based database)*. Disponible en <http://esa.un.org/MigFlows/Data%20availability.html>

MC: Datos Migración Colombia.

IBR: Inmigrantes, total por país de residencia anterior.

IFR: Inmigrantes extranjeros por país de residencia anterior.

IBC: Inmigrantes, total por nacionalidad.

jes sucesivos a Colombia. Es más frecuente por una parte de la colonia colombiana residente en Canadá, que las estancias no superen los doce meses dadas las facilidades de transporte aéreo, por lo que a efectos de la base de información escapan a la definición de emigrante que hemos implementado. Este es un problema que se ha planteado y que tiende a rebajar la estimación de emigrantes. El sobre-registro (133 por ciento) que presenta la estimación de flujos con destino Estados Unidos de Migración Colombia radica en que ésta mide flujos y los datos que proporcionan Naciones Unidas son inmigrantes colombianos que obtienen la residencia permanente en ese país. La propuesta sobre la que estamos trabajando consiste en evaluar un tiempo acumulado de residencia en el extranjero que permita soslayar la limitación de la definición unívoca de migrante internacional de Naciones Unidas. Este es, por lo tanto, uno de los desarrollos que pretendemos completar en el futuro.

En todo caso, los resultados son esperanzadores en la medida en que la estimación actual de los flujos emigratorios está por debajo de los niveles que puede alcanzar

Tabla 7. Nivel de cobertura de la emigración exterior colombiana según la base Migración Colombia. Promedio 2004-2009.

Países	Migración Colombia (MC)	Estadísticas países destino	% Cobertura MC
Alemania	1.428	1.738	82,2%
Australia	876	1.334	65,7%
Austria	155	87	178,2%
Bélgica	282	203	138,9%
Canadá	3.995	5.059	79,0%
Dinamarca	106	101	105,0%
España	25.871	26.316	98,3%
Estados Unidos	38.430	28.900	133,0%
Finlandia	49	35	140,0%
Francia	3.165	1.258	251,6%
Italia	2.415	2.138	113,0%
Israel	990	112	883,8%
Islandia	6	14	42,9%
Países Bajos	827	452	183,0%
Nueva Zelanda	43	37	116,2%
Noruega	169	151	111,9%
Suiza	586	495	118,4%
Suecia	382	280	136,4%
Total	79.774	68.710	116,1%

Fuente: Elaboración propia a partir de la base Migración Colombia y United Nations, *International Migration Flows to and from Selected Countries: The 2010 Revision (web-based database)* disponible en <http://esa.un.org/MigFlows/Data%20availability.html>

en el futuro la base Migración Colombia: en primer lugar, cuando se hayan recuperado las secuencias incoherentes y, en segundo lugar,

se incorpore una definición más flexible de emigración exterior que incorpore la duración de estancia acumulada fuera de Colombia.

Conclusiones

Tras la comparación de las estimaciones de emigración resultantes a partir de los métodos implementados en Migración Colombia sobre diferentes estadísticas españolas a partir de los microdatos de la Estadística de Variaciones Residenciales (EVR) y las recopiladas por Naciones Unidas podemos concluir que la estimación realizada es robusta por cuatro razones:

La primera razón es que, en el caso de España, en tres de las compa-

rativas la estimación se sitúa por debajo del valor estimado, pero en valores que oscilan entre el 68 y el 92%. No todos los individuos que viajaron a España que acabaron empadronándose pudieron equipararse a emigrantes por su corta estancia en España; otros llegaron a España vía otros países y una parte que desconocemos y que viajó a España no dejó huella estadística en el padrón. No existiría, por lo tanto, discordancia entre una y otra estadística.

La segunda es que los datos estimados por la base Migración Colombia recogen perfectamente la coyuntura migratoria registrada en España en los últimos años con una clara inflexión de los flujos en todas las tipologías para los años 2007-2008.

La tercera de las pruebas que validan este método reside en la transformación que sufre la estructura por edades de los viajes/viajeros derivada de la base Migración Colombia una vez aplicado el método

de estimación y convertidos en flujos migratorios¹⁴. Estos últimos muestran una elevada asociación entre la estimación de la estructura demográfica de los flujos emigratorios por edad, una vez sumada cada secuencia, y la estructura de entradas de colombianos a España por país de nacimiento y/o nacionalidad que recoge la Estadística de Variaciones Residenciales para el período 2004-2009 (gráfico 2).

Asimismo, los perfiles de emigraciones exteriores por sexo y edad para una serie de 30 países son coherentes con las características demográficas de los colombianos que residen en los países de destino (ver anexo gráfico), deducidas estas de fuentes censales y de encuestas, especialmente en lo

que atañe a la masculinización o feminización de la colonia de emigrantes y su especialización por edades y en consonancia con las estructuras demográficas propias de la migración, tal como muestran en sus diferentes trabajos Rogers y Castro (1982) y Rogers y Willekens (1986).

Finalmente, la comparación efectiva con otros destinos migratorios diferentes a los de España con las nuevas series publicadas por Naciones Unidas en la reciente *International Migration Flows to and from Selected Countries: the 2010 Revision (2012)* ha arrojado resultados dispares que pueden ser explicados desde la idiosincrasia de los sistemas estadísticos nacionales y también de algunas limitaciones de

los procedimientos implementados, cuya mejora pretendemos en futuro próximo.

En resumen, los positivos resultados obtenidos nos alientan a continuar en la depuración exhaustiva de las secuencias incoherentes y en especial en los procesos de imputación que supondrán una estimación al alza del flujo migratorio. En la hora de redacción de este trabajo se está procediendo a la extensión de los procesos estimativos al conjunto de secuencias y que pretendemos extender a un mayor número de países para, en fechas próximas, cotejar los resultados con los microdatos de la ronda censal de 2010.

Bibliografía

<<http://www.oim.org.co/anexos/documentos/publicaciones/libro128.pdf>> [20/04/2007]

Abel, G. J. (2010). Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 173(4), 797-825.

Cárdenas, M.; Medina, C. y Trejos, A. (2010). Measuring Economic and Social Impacts of Migration in Colombia: New evidence. *Borradores de Economía*, 601, Bogotá D. C.: Banco de la República.

Castro, L. J. y Rogers, A. (1982). What the age composition of mi-

grants can tell us? *Population Bulletin of the United Nations*, pp. 63-79.

Courgeau, D. (1973). Migrants et migrations, *Population* 28, pp. 95-129.

Cruz Zúñiga, P. G. (20 de octubre de 2007). Aproximación a los flujos y tipologías de la migración colombiana en España 1996-2006. *Biblio 3W Revista Bibliográfica de Geografía y Ciencias Sociales*, Universidad de Barcelona, 12(755). Disponible en <http://www.ub.es/geocrit/b3w-755.htm>.

Cruz, P. y Medina, R. (2006). La diáspora colombiana: derechos humanos y migración forzada Colombia-España (1995-2005). Construcciones sociales, políticas y jurídicas en el fenómeno de la migración forzada. Ponencia presentada en el 52.º Congreso Internacional de

Americanistas. Sevilla: 17-21 de julio de 2006.

Departamento Administrativo Nacional de Estadística, DANE (2007). Aproximación a la migración internacional en Colombia a partir del Censo General 2005. Recuperado el 18 de junio de 2006, de http://www.dane.gov.co/censo/files/presentaciones/migracion_internacional.pdf

— (2008). Estimación de la migración 1973-2005, Estudios post-censales número 6, DANE, Bogotá D. C.

— (2010). Informe final de evaluación de la calidad estadística del «Registro de entradas y salidas de personas del país», DANE-DIRPEN, Bogotá D. C.

De Beer J, Raymer J, Van der Erf R y Van Wissen L. (2010). Overcoming

¹⁴ Ver documento de trabajo Una nueva base de datos para la estimación de los flujos migratorios internacionales de Colombia: Metodología y resultados comparativos (2012), anexo gráficos 1 y 2. Disponible en: http://www.dane.gov.co/candane/index.php?option=com_content&task=section&id=18

the problems of inconsistent international migration data: A new method applied to flows in Europe. *European Journal of Population*, 26, 459-481.

Dumont, J. C. y Lemaitre, G. (2004). Counting Immigrants and Expatriates in OECD Countries: A New Perspective, Directorate for Employment Labour and Social Affairs, DELSA, OCDE.

International Organization for Migration, IOM (2004). *Return Migration Policies and Practices in Europe*. Geneva: IOM.

Kupiszewska, D. y Nowok, B. (2008). Comparability of statistics on international migration flows in the European Union. *International migration in Europe: Data, models and estimates*, Raymer J and Willekens F, eds., pp. 41-71. Chichester: Wiley.

Maguid, A. (2009). El estudio de la emigración internacional mediante los censos realizados en los países de origen: evaluación de resultados y recomendaciones. *Notas de Población*, 88.

Martí, M. y Ródenas, C. (2004). Migrantes y migraciones: de nuevo la divergencia en las fuentes estadísticas. *Estadística Española*, 46(156), 293-321.

Martínez, J. (2009). Medición e información sobre la migración internacional a partir de los censos: lecciones, desafíos y oportunidades. *Notas de Población*, 88.

Myers, D. (2004). Accuracy of Data Collected by the Census Question on Immigrants' year of arrival. Working Paper No. PDRG04-01.

Naciones Unidas (1999). *Recomendaciones sobre Estadísticas de las Migraciones Internacionales*. Revisión 1. Informes Estadísticos. Serie M, núm. 58, Rev. 1. Departamento de Asuntos Económicos y Sociales. División de Estadística. Naciones Unidas, Nueva York.

Organisation for Economic Co-operation and Development, OECD (2007). *International Migration Outlook*. Paris: OECD Publishing.

Ojeda, G. (2006). Una mirada de la familia y la migración: resultados del módulo migración internacional en la Encuesta de Demografía y Salud de 2005. En: Colombia nos une. Memorias II Seminario sobre migración internacional colombiana y la conformación de comunidades transnacionales, 15 y 16 de julio de 2006. [Libro en línea] Bogotá D. C.: Ministerio de Relaciones Exteriores de Colombia, pp. 53-57.

Poulain, M. (1993). Confrontation des statistiques de migration intra-Europe: vers une matrice complète? *European Journal of Population*, 9(4), 353-381.

— (1985). La migration. Concept et methodes de mesure. En Chaire Quetelet, *Migrations internes*, pp. 7-38.

Raymer, J.; De Beer, J. y Van der Erf, R. (2011). Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002-2007. *European Journal of Population*, 27(2), 185-215.

Raymer, J. (2007). The estimation of international migration flows: A general technique focused on

the origin-destination association structure. *Environment and Planning A*, 12, 371-388.

Recaño, J. (2011). *La emigración exterior de Colombia. Informe técnico del proyecto* núm. 1121. Bogotá D. C.: DANE.

Redstone, I. y Massey, D. (2003). Coming to Stay: An Analysis of the US Census Question Year of Arrival, paper presented at the Annual Meetings of the Population Association of America.

Rees, P., Bell, M. Duke-Williams, O. y Blake, M. (2000). Problems and solutions in the measurement of migration intensities: Australia and Britain compared. *Population Studies*, 54(2), 207-222.

Rogers, A. y Castro, L. J. (1982). Patrones modelo de migración. *Demografía y economía*, 16(3), 267-232.

Rogers, A. y Willekens, F.J. (Eds.) 1986. *Migration and settlement. A multiregional comparative study*. Dordrecht: Reidel Publ. Co.

Vaugelade, J. (1982). Stocks et flux dans l'analyse des migrants de retour. *Population*, 6, 1194-1199.

Willekens, F. (1985). Comparability of migration data: Utopia or reality? En Chaire Quetelet, *Migrations internes*.

Zlotnik, H. (1987). *Estimación sobre la emigración a partir de datos sobre residencia de hijos sobrevivientes: el caso de Colombia*. División de Población de Naciones Unidas.

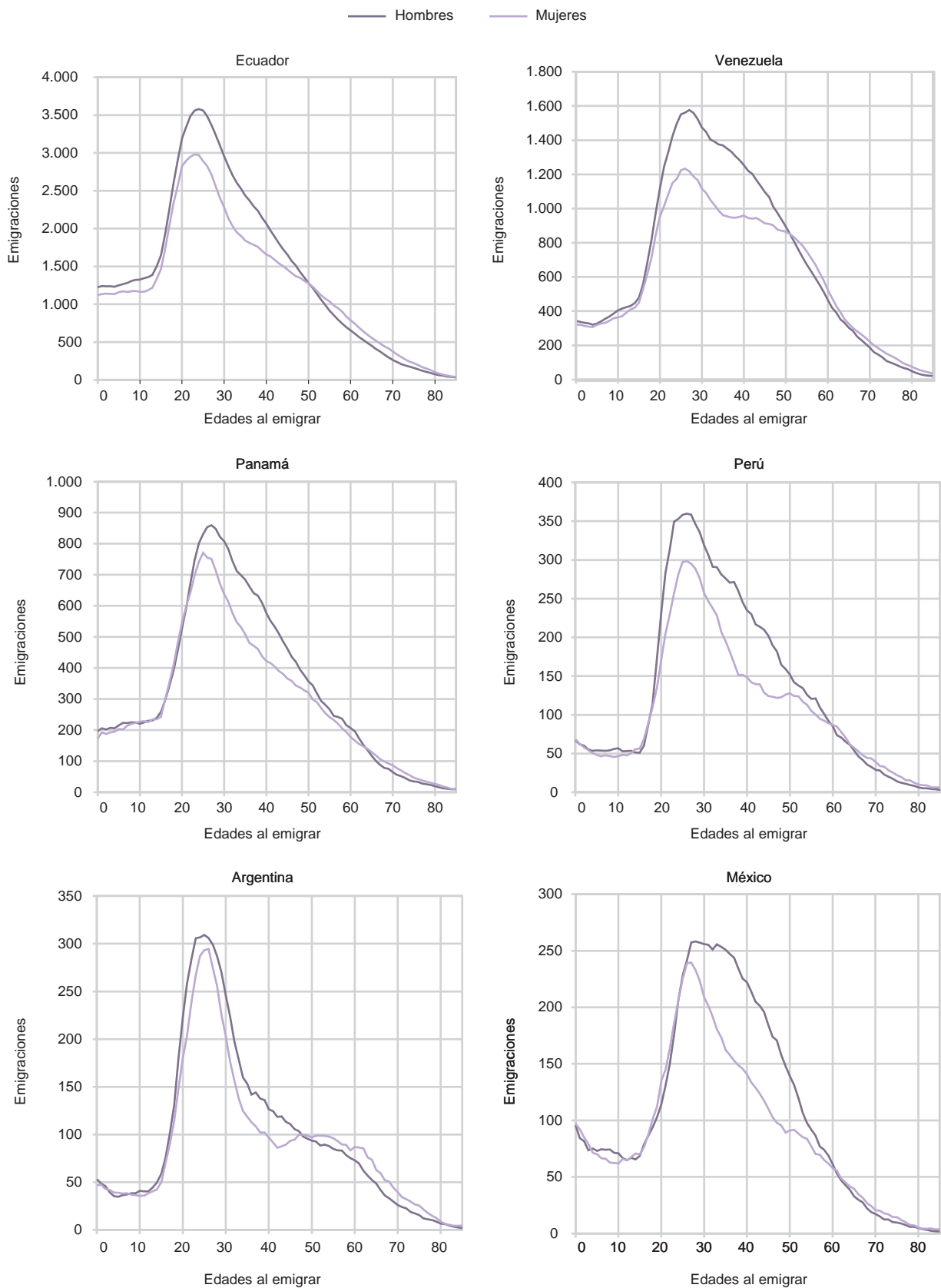
Anexos

Anexo 1. Estimación de flujos emigratorios según la base Migración Colombia y recopilación de flujos con origen Colombia en los países de destino según Naciones Unidas.

Base Migración Colombia	2004	2005	2006	2007	2008	2009
Alemania	2.206	1.060	1.025	1.388	1.342	1.545
Australia	626	472	686	1.076	1.273	1.120
Austria	171	88	126	125	172	245
Bélgica	334	186	188	236	326	422
Canadá	4.898	3.158	3.488	4.012	4.775	3.636
Dinamarca	115	75	90	101	117	137
España	27.290	18.031	25.147	32.846	30.231	21.680
Estados Unidos	52.788	32.014	31.620	33.226	39.018	41.911
Finlandia	47	53	41	52	53	48
Francia	3.333	2.197	2.015	2.904	4.020	4.518
Islandia	6	6	4	10	2	8
Israel	710	1.139	1.734	1.341	564	451
Italia	4.069	1.577	1.722	2.200	2.477	2.442
Noruega	256	144	138	171	164	140
Nueva Zelanda	48	36	30	38	49	57
Países Bajos	1.088	647	643	657	1.025	899
Suecia	527	295	311	351	457	350
Suiza	1.067	465	457	524	445	558
Flujos registrados en los países de destino	2004	2005	2006	2007	2008	2009
Alemania-IBR	1.576	1.547	1.752	1.654	1.896	2.001
Australia-IBR	490	680	1.160	1.990	2.350	
Austria-IFR	93	99	67	71	96	93
Bélgica-IFC	198	190	198	208		
Canadá-IFR	4.438	6.031	5.813	4.834	4.995	4.240
Dinamarca-IBR	71	93	114	94	126	109
España-IBR	17.413	21.351	28.650	36.434	34.577	19.469
Estados Unidos-IFR	18.055	24.705	42.017	32.055	29.349	27.221
Finlandia-IBR	35	35	39	32	37	31
Francia-IFC	1.124	1.169	1.556	1.221	1.221	
Islandia-IBR	13	36	11	8	5	11
Israel-IFR	53	135	142	232	55	52
Italia-IBR	2.673	2.042	1.878	1.856	2.242	
Noruega-IBR	164	145	163	173	152	108
Nueva Zelanda-IBR	25	40	23	33	49	52
Países Bajos-IBR	509	485	467	341	417	493
Suecia-IBR	240	248	345	273	299	273
Suiza-IFC	514	435	449	477	573	522

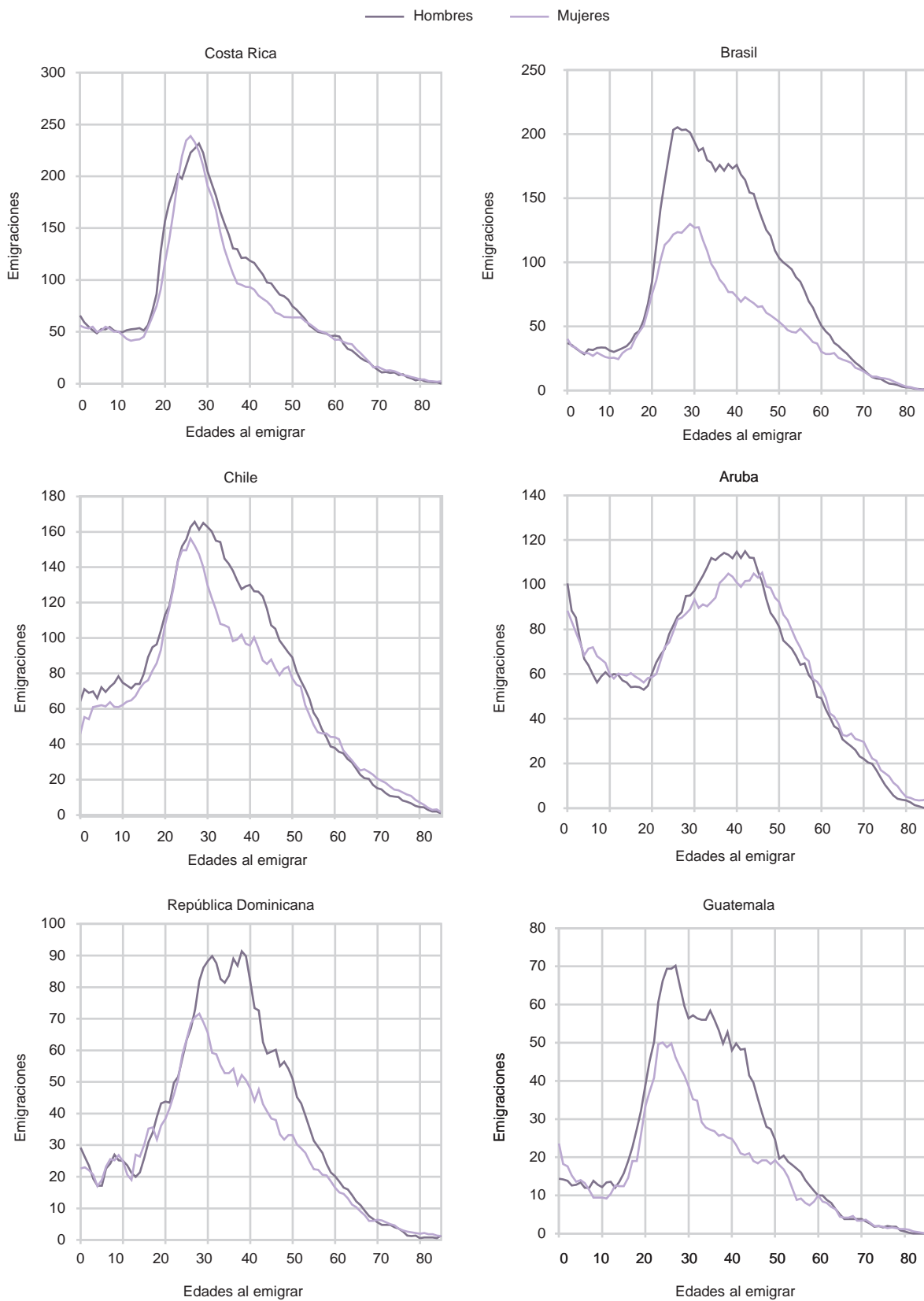
Fuente: Elaboración propia a partir de la base Migración Colombia y United Nations, *International Migration Flows to and from Selected Countries: The 2010 Revision (web-based database)*. Disponible en <http://esa.un.org/MigFlows/Data%20availability.html>

Anexo 2. Estimación de flujos de emigración exterior de Colombia (2004-2009). Total. Latinoamérica (I).



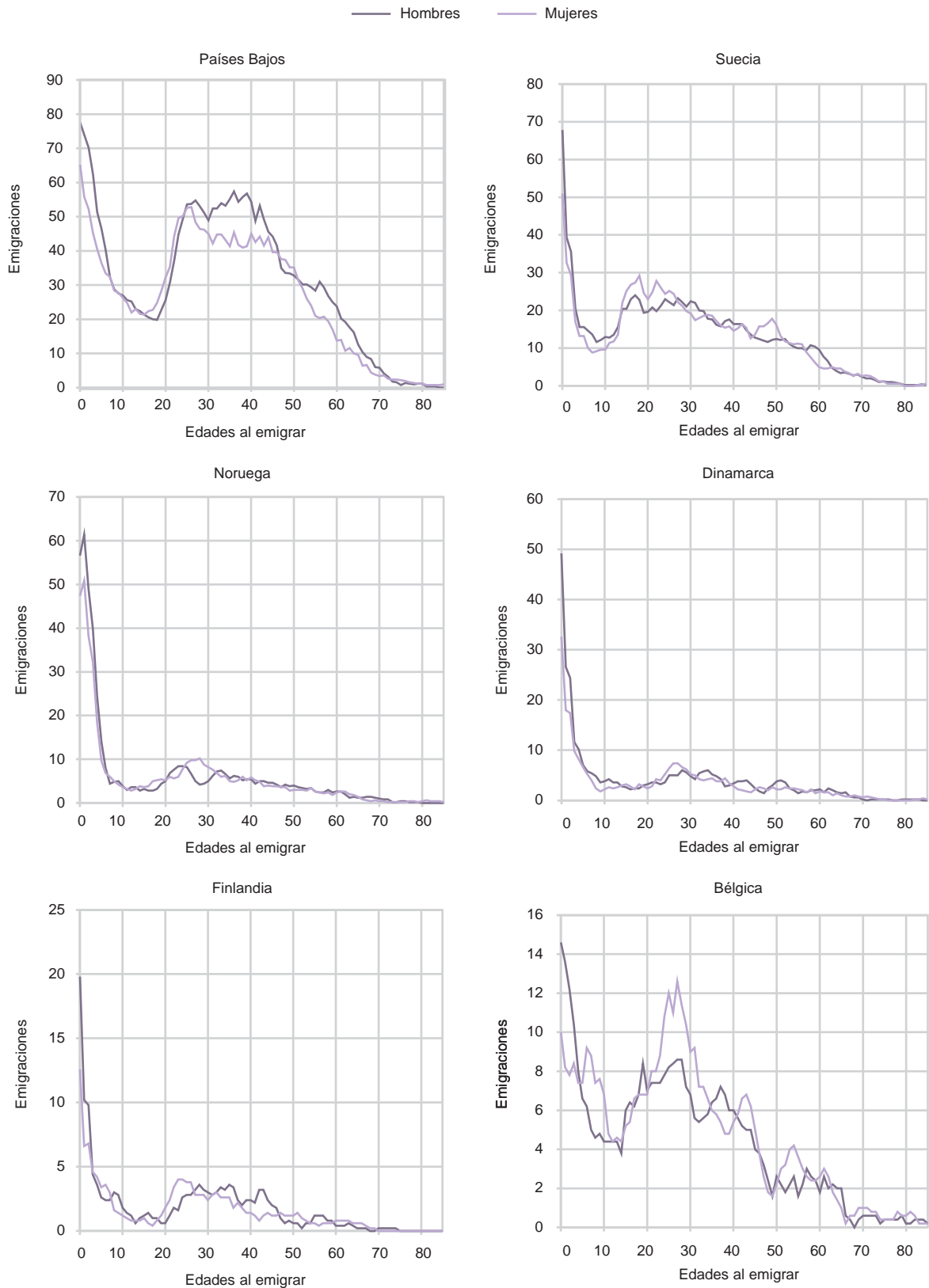
Fuente: Explotación DANE a partir de la base Migración Colombia. Media móvil de 5 términos. Elaboración propia.

Anexo 3. Estimación de flujos de emigración exterior de Colombia (2004-2009). Total. Latinoamérica (II).



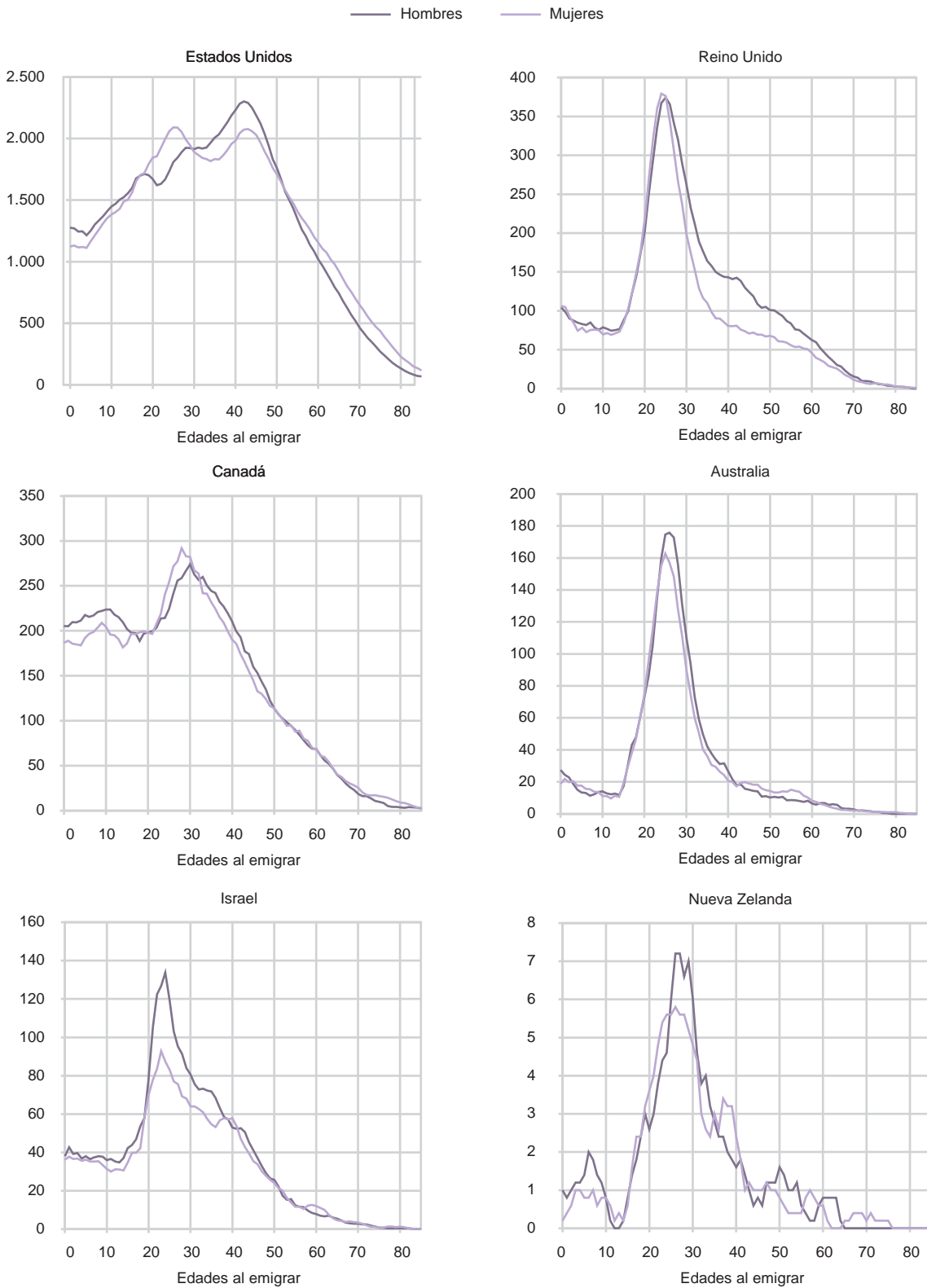
Fuente: Explotación DANE a partir de la base Migración Colombia. Media móvil de 5 términos. Elaboración propia.

Anexo 4. Estimación de flujos de emigración exterior de Colombia (2004-2009). Total. Modelo de los países nórdicos con flujo de adopciones.



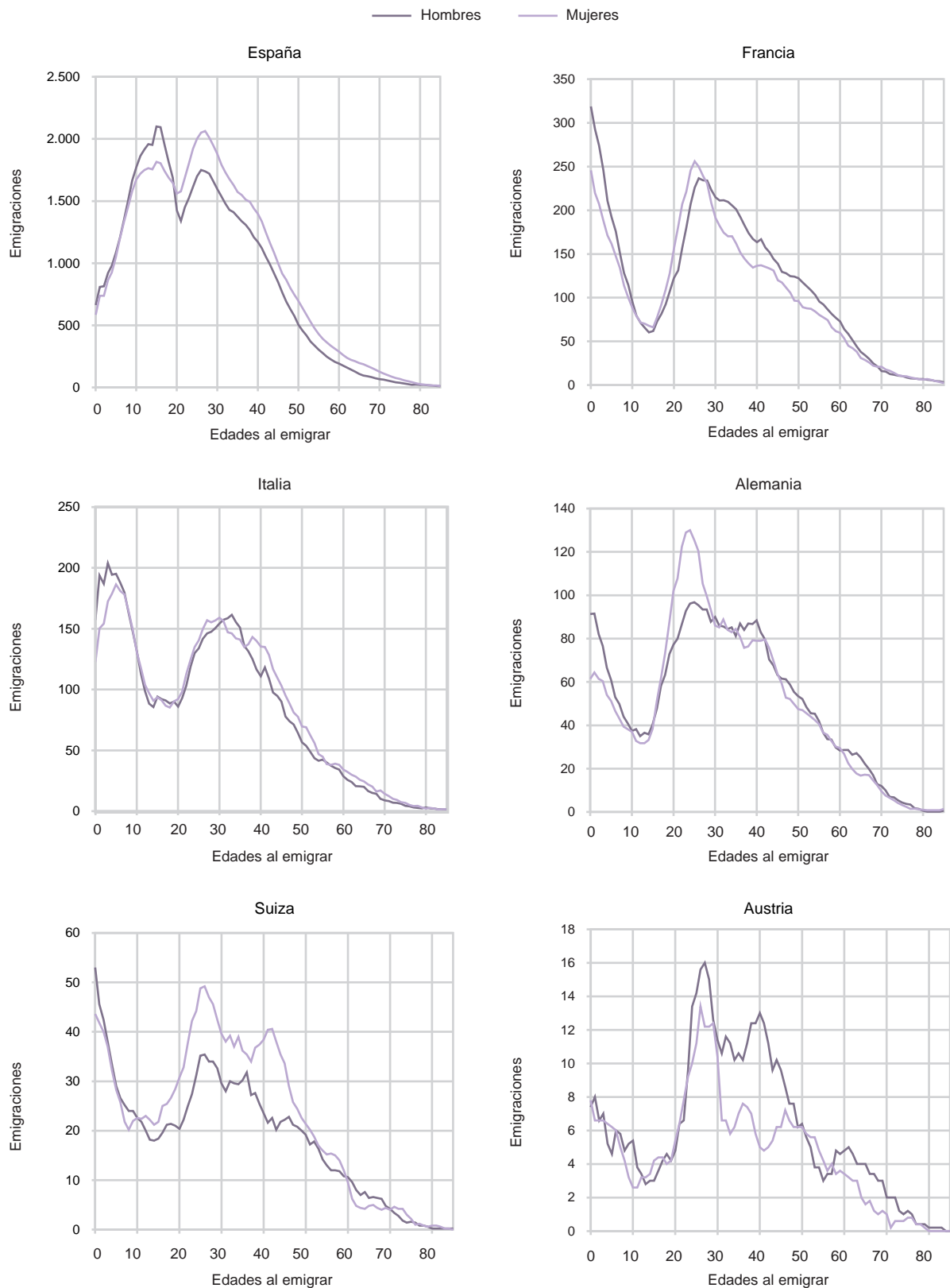
Fuente: Explotación DANE a partir de la base Migración Colombia. Media móvil de 5 términos. Elaboración propia.

Anexo 5. Estimación de flujos de emigración exterior de Colombia (2004-2009). Total. Estados Unidos, Gran Bretaña y otros.



Fuente: Explotación DANE a partir de la base Migración Colombia. Media móvil de 5 términos. Elaboración propia.

Anexo 6. Estimación de flujos de emigración exterior de Colombia (2004-2009). Total. Países de la Unión Europea.



Fuente: Explotación DANE a partir de la base Migración Colombia. Media móvil de 5 términos. Elaboración propia.

Compilación y síntesis de las metodologías internacionales aplicadas a procedimientos de retropolación

Sonia Milena Cifuentes Cruz
Andrés David García Ávila (q.e.p.d.)



Compilación y síntesis de las metodologías internacionales aplicadas a procedimientos de retropolación

Sonia Milena Cifuentes Cruz: Economista de la Universidad Católica de Colombia. Investigadora de la Dirección de Síntesis y Cuentas Nacionales del DANE.
Correo electrónico: smcifuentec@dane.gov.co

Andrés David García Ávila (q. e. p. d.): Economista de la Universidad Católica de Colombia, fue investigador de la Dirección de Síntesis y Cuentas Nacionales del DANE.

Fecha de recepción: 1 de octubre de 2012
Fecha de aceptación: 22 de agosto de 2013

Resumen: el creciente interés de los usuarios por disponer de series de tiempo históricas consistentes ha orientado los esfuerzos de los institutos de estadística hacia la búsqueda de métodos apropiados que garanticen continuidad en la información publicada. Este documento presenta las experiencias internacionales más sobresalientes en el empalme de series de tiempo, haciendo énfasis en el procedimiento, las variables empleadas y en algunos casos las limitaciones de los ejercicios desarrollados en cada país. La investigación permite realizar una propuesta aplicable a las series de agregados macroeconómicos en Colombia.

Palabras clave: año base, cuentas nacionales, empalme, encadenamiento, interpolación, retropolación, tasa de variación, SCN, SEC.

Abstract: the availability of consistent historical series has been an increasing interest of general users, which has lead the institutional efforts of statistical institutes towards the approach of appropriate methods that guarantee continuity in the information published. This paper contains the most outstanding international experiences on linking time series, making special emphasis on the estimation procedure, variables used and in some cases the limitations of the exercises developed in each country. This research allows developing an applicable proposal to macroeconomic aggregates time series in Colombia.

Key words: rebasing, national accounts, splicing backward, interpolation, retropolation, growth rates, SNA, ESA.

1. Introducción

La actualización del periodo base de referencia de las Cuentas Nacionales se lleva a cabo en cumplimiento de las recomendaciones internacionales (SCN93) y a fin de incorporar cambios metodológicos y estadísticos relacionados con la adopción de nuevos conceptos, clasificaciones, fuentes de información y procedimientos de cálculo. Tales cambios alteran la dimensión de los agregados macroeconómicos al impedir la comparabilidad entre las series publicadas para periodos anteriores, razón por la cual las entidades responsables de las estadísticas oficiales de un país deben revisar diferentes técnicas de empalme que les permitan generar series de tiempo largas y consistentes que sirvan como herramienta de análisis para académicos, investigadores e incluso hacedores de política económica.

Dentro de las alternativas metodológicas disponibles para realizar el empalme de series históricas se distinguen dos métodos: el Reproceso detallado y las Técnicas estadísticas de empalme, dentro de las que pueden situarse la Interpolación entre años base y el Método del indicador. El Reproceso detallado consiste en elaborar las cuentas nacionales para los años previos al año base, empleando los mismos conceptos, clasificaciones, fuentes y métodos del nuevo año de referencia. Aunque pareciera ser el método más apropiado, resulta poco factible debido a los requerimientos de información y recursos necesarios para su correcta aplicación.

Por su parte, la Interpolación considera que las cuentas nacionales se elaboran con el mayor detalle en los años base, razón por la que

estos se mantienen fijos durante el ejercicio y la diferencia originada entre los valores calculados con la base anterior y la nueva se distribuye mediante un promedio geométrico que permite estimar los periodos intermedios. Una de las alternativas del método del indicador es la Tasa de variación, que consiste en aplicar las tasas de crecimiento de la serie anterior, a los niveles obtenidos en el nuevo año base.

El objetivo de este documento es ofrecer un panorama general de los métodos usualmente empleados en diferentes países en el proceso de empalme de series de tiempo históricas, particularmente orientadas a la reconstrucción de las Cuentas Nacionales para el cálculo de los agregados macroeconómicos. La importancia de esta síntesis radica en que a partir del trabajo de investigación y estudio de las experiencias internacionales pueda definirse el método más apropiado para llevar a cabo la construcción de series largas, coherentes con los resultados del más reciente año base en Colombia; esto teniendo en cuenta algunas particularidades propias de la economía y el estado de la estadística básica disponible.

En la primera parte se exponen brevemente las experiencias internacionales en trabajos de retroposición, señalando en cada caso el periodo de tiempo considerado, las variables estudiadas, el método empleado y en algunos casos las inconsistencias que suelen presentarse en el empalme de series de tiempo. A continuación se presentan algunas conclusiones y resultados de la investigación, dentro de los que se encuentran algunas propuestas relacionadas con el encadenamiento de las cuentas nacionales en Colombia y su posterior aplicación.

2. Revisión y análisis de las metodologías internacionales asociadas al empalme de series de tiempo

A continuación se realiza un breve recuento de los trabajos de retroposición realizados por diferentes autores e institutos de estadística en diversas regiones del mundo. Para una mejor comprensión, los países se clasifican en dos categorías: desarrollados y subdesarrollados.

2.1. Métodos y alternativas empleadas por países desarrollados

Europa es sin duda uno de los mayores líderes en la producción estadística de series. Sus metodologías e investigaciones han servido de marco de referencia para muchos institutos y departamentos de estadística. En lo que se refiere a la aplicación de métodos de empalme de series históricas Europa se consolida como uno de sus pioneros, de hecho suelen desarrollar modelos macroeconómicos complejos como se describe a continuación.

Uno de los trabajos más relevantes fue elaborado por Ángel de la Fuente del Instituto de Análisis Económico (CSIC), para enlazar algunos agregados económicos nacionales y regionales en España para el periodo 1955-2007. Se comienza por enlazar la Contabilidad Regional de España en base 2000 (CRE2000) con series anteriores de la CRE y con datos de la Fundación BBVA. La serie homogénea elaborada por la Fundación BBVA (1999) cubre el periodo 1955-93 a intervalos bienales. Esta serie se extiende hasta 1995 en la Fundación BBVA (2000).

El INE ha elaborado cuatro series de Contabilidad Regional con bases 1980, 1986, 1995 y 2000. Las dos primeras y las dos últimas series han sido enlazadas por el INE para 1980-96 en base 1986 y 1995-2007 en base 2003.

Los principales problemas metodológicos que se presentan en la construcción de series enlazadas tienen que ver en primer lugar, con la elección de la serie de referencia en aquellos periodos donde hay varias disponibles. El procedimiento de selección consiste en analizar los determinantes de cada una de las variables con las que se está trabajando regresándola sobre alguna otra variable x_t , esto con el fin de determinar cuál de las series disponibles generará un estimador con mejores propiedades y con

un menor error cuadrático medio (ECM). El segundo problema es la selección del método de enlace que en particular intenta estimar y corregir el error de medición contenido en la serie más antigua.

En la construcción del enlace entre dos series generalmente suele presentarse una discrepancia entre la serie nueva (Y_t) y la serie antigua (X_t) en el momento de enlace (T). Dicha diferencia se expresa como:

$$d_T = \ln Y_T - \ln X_T = y_T - x_T$$

Y se debe únicamente a mejoras en la base estadística que permite una medición más apropiada de algunas actividades o sectores que previamente no estaban cubiertos de manera adecuada (afloramiento de nueva actividad).

(año base). Aquí se supone que el error de medición de la serie más antigua se ha generado entre 0 y T , a un ritmo uniforme. De esta manera, debería sumarle a la serie antigua una fracción creciente de d_T ¹.

$$\hat{y}_t^i = x_t + \frac{t}{T}(y_T - x_T) \equiv x_t + \frac{t}{T}d_T$$

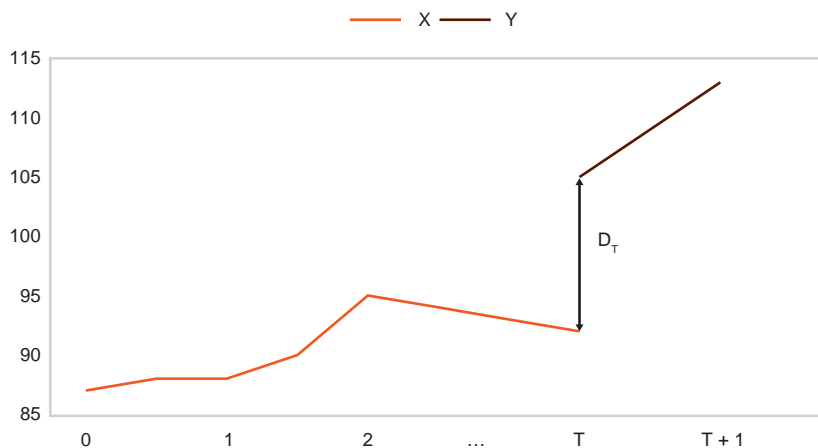
Debido a que los procedimientos de enlace habituales imponen hipótesis fuertes acerca del comportamiento de este error, se desarrolla un procedimiento mixto que intenta parametrizar el error de medición inicial y hacer que dependa del ritmo relativo de crecimiento de las actividades en relación con el resto de la economía.

$$\hat{y}_t^m = x_t + \hat{d}_t^m \text{ para } t \leq T \text{ con } \hat{d}_t^m = d_T \rho^{\frac{T-t}{T}}$$

La manera más sencilla de construir series enlazadas de magnitudes reales consiste en enlazar primero las magnitudes medidas a precios corrientes por el procedimiento más oportuno, para luego enlazar por retropolación los correspondientes índices de precios y por último, deflactar la serie nominal construida inicialmente utilizando esta serie de precios. Alternativamente, se pueden enlazar directamente las series a precios constantes y luego recuperar la serie enlazada de deflatores como la relación entre la serie enlazada nominal y la correspondiente magnitud real.

Un estudio similar realizado por el INE de España consiste en obtener una serie homogénea del PIB regional y del Valor Añadido Bruto por ramas de actividad estimados a precios corrientes y en términos de volumen encadenado, así como del empleo total y asalariado en términos de puestos de trabajo para el periodo comprendido entre 1995 y 2006. La desagregación propuesta por rama de actividad es la siguiente: Agricultura, ganadería y pesca; Energía; Industria; Construcción; Servicios y Servicios de no mercado.

Gráfico 1. Discrepancia entre la serie nueva y la serie antigua en el momento del enlace.



Fuente: Fuente (2009).

Los métodos de enlace más comunes son la retropolación y la interpolación entre años base. El primero consiste en extender hacia atrás la serie nueva usando las tasas de crecimiento de la serie antigua. De esta manera la retropolación de Y_t tomando como referencia X_t está dada por:

$$\hat{y}_t^r = x_t + (y_T - x_T) \equiv x_t + d_T \text{ para } t \leq T$$

La hipótesis en este caso es que el error de medición en la serie antigua es constante, es decir que ya existía en el momento 0. Así, para obtener el valor correcto de la magnitud de interés solo debe añadirse a la serie antigua la discrepancia observada en el momento del enlace (d_T).

En el caso de la interpolación, el enlace se realiza forzando la extensión de la nueva serie a pasar por cierto punto de la serie antigua

¹ En este caso se conserva el valor original de la serie antigua en su año base, pero no su tasa de crecimiento.

La base 1995 de la Contabilidad Regional de España (CRE) introdujo cambios metodológicos importantes, pues fue el resultado de la implantación del Sistema Europeo de Cuentas Nacionales y Regionales 1995 (SEC-1995). Por su parte, la base 2000 incorpora en la contabilidad nacional y regional cambios tanto conceptuales como estadísticos. Dentro de los cambios conceptuales se encuentra el tratamiento de los Servicios de Intermediación Financiera Medidos Indirectamente (SIFMI), pues aparte de establecer una manera para calcularlos, deben asignarse a los sectores y ramas de actividad que los usen. Los cambios estadísticos básicamente están sustentados en la introducción de nuevas fuentes de información estadística, que surgen desde la base anterior.

La CNE-2000 y la CRE-2000 determinan que el nivel de los agregados a precios constantes no va a estar referido a un año base fijo, sino que la referencia va a ser siempre el año precedente (base móvil). El inconveniente que presenta este método es la falta de aditividad de los datos. A su vez, el principio que rige el enlace de las series en ambas bases es que las estimaciones de los años de referencia tienen carácter estructural, y por lo tanto, desde un punto de vista estadístico, no van a ser revisadas. De esta manera, las únicas diferencias en las estimaciones del año 1995, bases 1995-2000, serán las relacionadas con cambios metodológicos.

El primer paso para enlazar las series del valor añadido a precios corrientes es la estimación de las diferencias metodológicas que, a precios corrientes, solo afectan el tratamiento de los SIFMI. De esta manera, debe estimarse la cantidad de SIFMI que es consumo intermedio de cada rama de

actividad para cada territorio, e igualmente, será necesario revisar la producción de la rama de actividad intermediación financiera, para cada territorio. Tras eliminar el efecto SIFMI, las únicas diferencias que existen en los datos de las dos bases se explican por cambios estadísticos. Así, en el año 2000 existirán dos estimaciones: la primera obtenida por extrapolación de los datos con estructura de la base 2005; y la segunda, estimada directamente en el marco de la base 2000. Debido a que no se cuenta con información de cómo y en qué momento se han generado las diferencias entre ambas estimaciones, se supone que lo han hecho progresivamente a lo largo del tiempo, desde el inicio de la base anterior.

Suponiendo que X_{95} , X_{96} , ..., X_{00} son las estimaciones del valor añadido a precios corrientes de una determinada rama de actividad, base 1995, una vez considerado el efecto SIFMI, Y_t será la estimación del mismo valor añadido en base 2000. Luego, el cociente entre Y_{00} y X_{00} es la desviación que han supuesto los cambios estadísticos en dicho valor añadido.

Si se supone que esta desviación afecta al crecimiento interanual de manera progresiva y acumulativa, desde 1995 hasta 2000 (5 años), se tiene que:

$$d = Y_{00} / X_{00}$$

Repartiendo entre los 5 años, el efecto anual será:

$$r = d^{1/5} = (Y_{00} / X_{00})^{1/5}$$

Por lo tanto, los valores en la base 2000 serán:

$$Y_{95} = X_{95}$$

$$Y_{96} = X_{96} * r$$

$$Y_{97} = X_{97} * r^2$$

$$Y_{98} = X_{98} * r^3$$

$$Y_{99} = X_{99} * r^4$$

$$Y_{00} = X_{00} * r^5$$

Las diferencias entre el dato nacional y la agregación de los territoriales obtenidos por el procedimiento descrito serán repartidas de forma proporcional al peso de cada territorio dentro de las diferentes ramas de actividad. La estimación del PIB regional se obtendrá añadiendo al valor añadido calculado, los impuestos sobre los productos menos subvenciones. La asignación de dichos impuestos netos consiste en repartir su montante nacional de acuerdo con el peso del valor añadido bruto de cada territorio². Dado que la base 2000 no introdujo cambio metodológico alguno sobre la Remuneración de Asalariados, las diferencias entre los valores territoriales de las bases 1995 y 2000 se explicarán por razones estadísticas. Además sobre este agregado se aplicará el mismo procedimiento utilizado para corregir el VAB.

Respecto a la valoración a precios constantes del año anterior e índices encadenados, se tiene que en la base 2000 la estimación del crecimiento en volumen de un agregado entre dos años consecutivos, t-1 y t, se realiza valorando dicho agregado en t con la estructura de precios de t-1. Los índices de precios empleados corresponden a los que ya se disponían en la base 1995.

Francia es uno de los países precursores en la aplicación de métodos de retropolación, liderados, entre otros, por el Instituto Nacional de Estadística y Estudios Económicos (INSEE). El primer trabajo presenta un método de retropolación de las cuentas nacionales, aplicable a las cuentas de bienes y servicios, cuentas de los sectores institucionales o cuentas financieras. El método consiste en mantener como evoluciones de la nueva base (NB) las evoluciones modificadas de la

² Para la estimación del PIB a precios constantes del año anterior y la asignación de los impuestos netos, se aplica un criterio similar.

base anterior (AB). Así, si se dispone de dos conjuntos AB y NB, que se superponen en al menos un punto en el tiempo, debe trabajarse serie por serie al grado de detalle elegido y para cada serie NB, se busca o se reconstruye de manera simple su equivalente en AB corregida de los posibles errores, modificaciones en conceptos y/o nomenclaturas. Posteriormente se recalculan los totales de empleos y recursos por productos u operaciones y finalmente se equilibran las diferencias que puedan presentarse.

El método empleado se desarrolla en varias etapas. En la primera se decide la amplitud de la retropolación. En la segunda, se crea la «estructura informática», donde a partir del archivo que contiene los datos de NB, se crean dos más, uno para almacenar los conceptos y la nomenclatura de AB y otro para organizar la antigua base modificada (ABm), manteniendo la estructura de NB. En la tercera etapa, se calcula ABm examinando serie por serie los cambios estadísticos y metodológicos que deben incluirse en AB. Uno de los aspectos más delicados en este punto es la aparición del IVA y la distribución de los SIFMI. Tras haber realizado esto, para cada serie de NB debe existir una serie de ABm equivalente.

Partiendo del principio que señala que las evoluciones entre ABm y NB son idénticas, en la última etapa se lleva a cabo la retropolación serie por serie, según la fórmula:

$$NB_{(t-1)} = NB_{(t)} \times ABm_{(t-1)} / ABm_{(t)}$$

Esta ecuación se aplica durante todo el periodo de tiempo a partir del primer punto conocido de NB.

Dado que se retropolan los distintos elementos de manera independien-

te, se hace necesario equilibrar, de forma tal que Total usos = Total empleos. Al respecto, el método sugerido propone distribuir la diferencia proporcionalmente sobre el conjunto. Por ejemplo si $\Sigma R > \Sigma E$, se va a buscar $k > 0$ tal que $R' = (1 - k) * R$ y $E' = (1 + k) * E$ con $\Sigma R' = \Sigma E'$. Es decir:

$$k = (\Sigma R - \Sigma E) / (\Sigma R + \Sigma E)$$

Este cálculo supone que R y E son positivos, pero en realidad algunos recursos pueden ser negativos (márgenes, subvenciones) y en el caso de los empleos, la variación de existencias no siempre tiene el mismo signo. Se hace necesario entonces modificar el cálculo anterior.

Se sustituye

$$R' = (1 - k) * R$$

por

$$R' = (1 - k * \text{sign}R) * R \\ R = (1 - k * \text{abs}R / R) * R' \\ R = R - k * \text{abs}R$$

y simétricamente por $E' = E + k * \text{abs}E$, donde

y en consecuencia $|k'| < |k|$.

Es importante tener en cuenta que no es necesario reequilibrar todas las variables y que las etapas 4 y 5 pueden realizarse de dos maneras. Se retropolan todas las series elementales independientemente y se equilibra simultáneamente durante el periodo o se retropolan y se equilibran año por año.

Los cálculos anteriores se realizan para los valores corrientes; la etapa 6 consiste en estimar los volúmenes (a precios del año anterior), dividiendo estos valores por los índices de precios de la antigua base. El resultado son series NB en volumen no equilibradas, que deben conciliarse.

La última etapa es la agregación de las cuentas en valor y en volumen

que se hace por simple adición³. Aquí es necesario comparar las evoluciones de los agregados globales en AB y NB retropolados, que deben ser cercanos por construcción.

La siguiente propuesta del INSEE en la construcción de series de tiempo se realiza para el periodo 1949-1985, base 1970. El documento comienza por señalar las fuentes de información para las series 59-85, base 70, y 49-59 de los 80. Posteriormente se menciona la manera en que se almacenaron los datos y luego la nomenclatura de los productos de las bases 70 y 00. En este punto se resaltan las correcciones realizadas sobre los bienes y servicios en la base 70 de acuerdo con los cambios incorporados en la nueva base. Posteriormente se realizan algunas observaciones acerca de la valorización de las operaciones de bienes y servicios a precios corrientes y constantes de 1970; también la manera de medir la producción, ya que se hace necesario expresar todo en términos de precios productor. Al respecto se menciona la dificultad de reconstruir los impuestos y las subvenciones en base 70; y a falta de información, se opta por evolucionar producto por producto, a partir del año 59 de la base 2000.

En cuanto al SIFMI se tiene que una vez conocido el total en la retro 59, este se distribuye en cuotas relativas por operaciones y sectores, según la evolución del tipo de mercado monetario. En el caso de los márgenes de transporte, se aplica el mismo método empleado en la retro 59; es decir, se mantiene la evolución en volumen del producto en consideración y la evolución en precio de la producción de transporte.

Los índices de volumen para n-1 son construidos según la siguiente fórmula:

$$VOL_{(n,b00)} = \frac{VAL_{(n-1,b00)} \times VOLPCH_{(n,b70)}}{VOLPCH_{(n-1,b70)}}$$

³ Las cuentas en volumen encadenadas se obtienen por el método habitual.

Tras realizar estos trabajos hay que garantizar el equilibrio entre los recursos y los empleos, realizando una distribución proporcional sobre el conjunto de productos.

En el documento relacionado con la retropolación anual de las cuentas nacionales base 2000, se realiza una revisión minuciosa de la serie 1978-2003, base 95, y a su vez de la serie 1959-1997, base 80, con el fin de realizar un conjunto de modificaciones sobre el periodo 1978-1999, que a su vez permitan dar un orden coherente a los datos de las cuentas nacionales e incorporar algunas modificaciones a las tasas de crecimiento del PIB ligeramente en 1990 y más sustancialmente en la década de 1980. De esta manera, las tasas de crecimiento guardarían una mayor coherencia con la base 80.

Los parámetros empleados para determinar los cambios que deberían aplicarse a las series originales fueron:

- Que la diferencia entre la evolución de la base 2000 sea bastante amplia respecto a las evoluciones de las bases precedentes.
- El criterio para incorporar los cambios metodológicos en la serie 1978-1999 a precios corrientes fue que las cifras consideradas superaran el millón de euros. La gran mayoría de estos cambios se dieron para variables como la inversión y la variación de existencias, así como los reequilibrios se aplicaron sobre la producción y el consumo intermedio.

El documento hace especial énfasis en la necesidad de incorporar cuatro innovaciones conceptuales relevantes al relacionar la base 80 con la 2000:

a. *La valoración de la producción a precios básicos:*

$$P1PP = P1 + D214 + D31 + P91$$

Donde:

P1PP: *Producción a precios productor*

P1: *Producción a precios básicos*

D214: *Impuestos sobre los productos*

D31: *Subvenciones*

P91: *Márgenes de transporte*

En la base 80 la producción era valorada a precios productor (P1PP). La diferencia entre ambas observaciones radica principalmente en los otros impuestos sobre los productos (D214), sobre las subvenciones a los productos (D214) y sobre los márgenes de transporte (P91).

Fue necesario realizar un trabajo muy específico en esta retropolación para hallar los impuestos y las subvenciones, ya que estos nunca fueron publicados a nivel de producto en la base 80. Diferentes conceptos tributarios (TIPP, los derechos sobre el alcohol y el tabaco, los impuestos sobre los seguros, los derechos de registro) han sido trasladados para reconstruir la serie de impuestos a los productos. Para retropolar las subvenciones antes de 1978, se utilizó la misma relación empírica que se utiliza sobre el periodo 1978-1997 teniendo presentes las diferentes subvenciones catalogadas en la base 80 y las subvenciones de los productos en base 95 y 2000. En cuanto a los márgenes de transporte, estos son una innovación de la base 95.

b. *Introducción de los márgenes de transporte a partir de la base 1995:* Esto permite estimar con una mayor precisión la producción, separando la actividad secundaria del transporte. En la retropolación, estos márgenes son estimados en constantes, haciendo que evolucionen tal

como lo hace el uso total de un producto, y el corriente, utilizando el precio de la producción del transporte.

c. *Gasto individualizable en investigación y desarrollo (FN4), APU (FR1) y actividades de asociación (FR2):* Para las administraciones públicas, el gasto de consumo final individualizable comprende la atención colectiva de ciertos gastos de los hogares (prestaciones en especie de seguros de enfermedad, ayudas oficiales a la vivienda, reducciones sobre los transportes públicos, principalmente) que fueron descritos en base 80 como prestaciones sociales. Las series de gastos de consumo individualizables ya han sido calculadas y publicadas hasta 1959. Excepto para la salud de no mercado que prácticamente ha sido replicada sin ajustes en el marco de esta retropolación. Para los tres productos en los que existe un gasto en consumo colectivo de las administraciones públicas (investigación y desarrollo [FN4], la administración pública [FR1], actividades de la asociación [FR2]), este agregado se obtuvo por diferencia entre el gasto total de consumo, con el retropolado de la base 80, y los gastos de consumo individual.

Distribución de los SIFMI: La primera modificación introducida por la base 95 es una consolidación parcial de la producción y el consumo intermedio interbancario. Una segunda innovación introducida por la base 2000 fue la asignación de los SIFMI entre los distintos sectores institucionales de la economía. Por último, la base 2000 exige que se realicen las cuentas de consumo final y comercio exterior de los SIFMI, ya que anteriormente, hasta la base 95, eran considerados como consumos intermedios. Desde 1978, el cálculo de estas

operaciones se basa directamente en los datos de créditos incurridos y los depósitos sobre los créditos correspondientes (márgenes). Sin embargo, estos datos no existen antes de 1977. La retropolación 1959-1977 se basa en la evolución de la tasa del mercado monetario, cuyos nexos con las partes relativas de las diferentes correcciones de intereses en la producción total de SIFMI, han sido calculados sobre el período 1978-2004. La utilización de los indicadores de evolución de la base 80, hicieron posible prolongar esta retropolación en base 2000 hasta 1959 para el PIB y sus componentes.

En algunos trabajos aplicados en la Eurozona una de las alternativas consiste en realizar una retropolación condicionada con datos de alta frecuencia, empleando series relacionadas. Para ello se recurre al modelo estático de regresión de Chow y Lin y se considera una simple extensión dinámica para estimar series encadenadas e indicadores que den mayor flexibilidad a la propuesta. El procedimiento consiste básicamente en estimar los valores trimestrales perdidos en el pasado, para lo cual se dispone de *benchmarks* anuales, series trimestrales recientes y una o más series relacionadas que cubren los trimestres que se desean retropolación.

El uso de modelos econométricos está muy sujeto al cumplimiento de supuestos y restricciones de acuerdo con características muy puntuales de la región objeto de estudio, por lo que podría considerarse que de cierta manera pueden tener limitaciones al ser empleados en casos más generales. Por ejemplo, para encontrar un indicador de retropolación del Euro 7, que contiene mayor volumen de información, y no del

Euro 15 (debido a las limitaciones en la disponibilidad de datos), se adopta un método mixto en el cual se realiza una regresión aplicando un modelo ARIMA. Inicialmente se consideró un modelo lineal que luego se convertiría en un modelo estructural, donde la descomposición de la serie de tiempo se hace con un retroceso detallado para después retropolación con el indicador previamente calculado a nivel trimestral en una mezcla de métodos econométricos con *benchmark*.

Si las series están relacionadas, se recomienda cointegrar el indicador a la serie para producir mejores resultados (supone tendencias comunes en las series) y hacer una revisión de sus propiedades estadísticas.

Aunque en el caso de Canadá no fue posible tener acceso a metodologías específicas de retropolación, se recogieron algunas experiencias que proporcionan una visión cercana al encadenamiento de series en este país. En primera instancia se tiene que el cálculo del PIB mensual basado en la producción consiste en la sumatoria de los valores agregados generados por las industrias. Esta práctica se lleva a cabo con el fin de evitar la doble contabilidad de aquellos productos producidos por una industria y usados como consumo intermedio de otras, durante el mismo mes. El objetivo que persigue el cálculo del PIB desde esta óptica es mostrar cómo cambia el valor de la producción con el paso del tiempo y describir tales cambios sin que sean afectados por las fluctuaciones de precios, generando así estimaciones que reflejen variaciones solo en volumen.

Hay algunas alternativas acerca de los precios que pueden emplearse en la comparación de las cantidades producidas en diferentes periodos del tiempo. El primer tipo de valoración produce la llamada me-

didada de volumen tipo Laspeyres y se expresa como:

$$V_{\text{Laspeyres}} = \frac{\sum P_{t-1}q_t}{\sum P_{t-1}q_{t-1}}$$

Donde q es la cantidad y p es el precio, t es el periodo actual y 0 denota el periodo referencia (periodo base). La sumatoria cubre todas las mercancías producidas por una industria en particular. Dado que en este tipo de cálculo se aplican precios fijos, el método Laspeyres permite la comparación entre el volumen de producción en dos periodos durante el lapso de tiempo cubierto.

En cambio el procedimiento mediante el cual los precios actuales son usados para valorar las cantidades producidas en un periodo anterior, recibe el nombre de medida de volumen tipo Paasche y está dado por:

$$V_{\text{Paasche}} = \frac{\sum P_t q_t}{\sum P_t q_0}$$

El método de Paasche tiende a distorsionar las comparaciones entre periodos, excepto las del periodo actual y el periodo elegido como base.

En la construcción de las medidas de volumen tipo Laspeyres y Paasche, se supone que el precio de las mercancías no cambia significativamente durante los periodos que han sido comparados, o que si lo hacen, el cambio se origina en la misma proporción. Inicialmente el método empleado en Canadá para estimar el volumen de la actividad económica era el Laspeyres. Al respecto debe anotarse que el uso de los precios de un año base fijo a lo largo de un periodo de tiempo prolongado produce sesgos importantes cuando la estructura de precios del periodo actual llega a ser muy diferente a la que existía en el año base⁴.

Para superar la discontinuidad generada por el cambio de año base en las series del PIB mensual, los precios del nuevo año base se apli-

⁴ En Canadá el cambio de año base se realizaba con un intervalo de diez años (1961, 1971 y 1981). Luego de 1981 el intervalo se redujo a 1968, 1992 y 1997, año a partir del cual el cambio de año base se realiza con una frecuencia anual. La información más actualizada del PIB mensual se expresa en términos de precios de 2003.

can solo hacia adelante y para periodos anteriores al año de referencia, la serie de precios constantes en cada nivel de detalle se ajusta al nuevo año base por segmento multiplicándola por un factor de enlace que simplemente es la relación en el nuevo año base entre el PIB valorado a los nuevos precios y a los del año base anterior. El proceso anterior es llamado encadenamiento y su objetivo es eliminar los rompimientos entre años base consecutivos preservando el comportamiento en volumen a cada nivel de agregación. Con este método, las tasas de crecimiento originales solo pueden mantenerse a costa de perder la aditividad.

Dado que Canadá dispone de Cuentas Insumo-Producto que comprenden no solo estimaciones del factor ingreso y la demanda final por producto, sino también estimaciones del Valor Agregado por industria y producción por producto, surge la inquietud entre los asesores acerca de la posibilidad de encontrar en Colombia cuentas semejantes que puedan utilizarse como años de referencia para la serie 1975-2005. Se propone además una alternativa para suplir la información faltante, que consiste en apoyarse en datos de encuestas o, en su defecto, indicadores líderes que permitan estimar los años intermedios. Al respecto es importante tener en cuenta que: «Una retropolación entre dos puntos conocidos puede realizarse i) con algunos ajustes de juicio a los datos a fin de tener en cuenta algunos eventos económicos específicos o problemas con las fuentes de información y ii) basándose en indicadores asociados o proyectores de diferente naturaleza dependiendo de las fuentes de datos disponibles»⁵.

El estudio de la literatura también se orientó a la revisión de dife-

rentes documentos del continente asiático. Aunque algunos de ellos no exponen un método definido de empalme de series, proporcionan una visión general de la manera como abordan diferentes aspectos alrededor de su sistema de cuentas nacionales.

Desde 1984, los institutos de estadística han optado por revisar el sistema de contabilidad original (MPS) con el fin de satisfacer las necesidades que han surgido a partir de la implementación de reformas económicas. En China, por ejemplo, hasta 1992 el Sistema de Cuentas Nacionales siguió los principios del SNA 1968 y el MPS fue establecido. Algunas sugerencias del proyecto del SCN 1993 fueron adoptadas en el nuevo sistema, que combina el SNA y el MPS. Tras poner en práctica este nuevo sistema se cancelaron los contenidos relacionados con el MPS en 1994, año en el cual el SNA se convirtió en la única guía de la contabilidad nacional. La nueva versión del Sistema de Cuentas Nacionales de China (2000 CSNA) está principalmente inspirada en los principios del SCN 1993.

Aunque el SCN de China dispone de cinco años base (1952, 1957, 1970, 1980 y 1990), aún no se realiza el enlace de las series del PIB. El NBS (National Bureau of Statistics, P. R. China) propone enlazar las series de tiempo del PIB a precios constantes, no solo para el nuevo año base sino también para el año base previo.

Hong Kong, por su parte, compila el PIB basado en el gasto a precios constantes. El año de referencia actual a precios constantes de la serie del PIB basado en el gasto es 1990. Los años base anteriores fueron 1966, 1973 y 1980. Durante la última sesión de reajuste en la que el año base se ha actualizado desde 1980 hasta 1990, las series

a precios constantes de 1986 en adelante se obtienen reponderando los precios promedio en el nuevo año base, 1990.

La metodología general implica la compilación de los índices de precios reponderados para deflactar los componentes del gasto, y la adición de los componentes deflactados para obtener agregados a precios constantes en el nuevo año base. Para la serie de datos anteriores a 1986, las cifras a precios constantes compiladas en el año base anterior tanto a nivel agregado como por componentes se enlazaron a las series reponderadas por encadenamiento simple con 1986 como el año de enlace. Esto con el fin de producir una serie comparable a precios constantes basados en el nuevo año base de 1990.

Las series del PIB enlazadas antes de 1986 mantienen las mismas tasas de crecimiento real antes y después del enlace. Durante el último cambio de base se incorporan a la serie del PIB nuevas fuentes de datos, métodos de estimación y algunos de los nuevos lineamientos del SNA 1993. Como resultado de estas inclusiones las tasas de crecimiento del PIB basado en el gasto y sus componentes, antes y después del ejercicio de enlace, son diferentes.

India ha realizado el ejercicio de empalme para cuatro periodos, revisando los años base de 1948-49 a 1960-61, 1970-71, 1980-81 y el más reciente 1993-94. En todas las ocasiones el empalme de la serie actual se ha llevado hasta 1950-51, formando una serie coherente. En cuanto al método de enlace, se intenta representar los cambios de cobertura de acuerdo con la disponibilidad de información en las bases precedentes. En ausencia de información, se adopta la técnica de empalme o se mantiene la tasa de crecimiento observada en la serie anterior.

⁵ Este comentario fue realizado por el consultor Michel Séruzier, para efectos de los trabajos de retropolación de las Cuentas Nacionales de Colombia.

Las series de Cuentas Nacionales de Corea se han actualizado en ocho ocasiones; el primer cambio de base ocurrió en 1964 y el último se llevó a cabo en 1999. En Corea el año base se reajusta cada cinco años. El ejercicio de empalme en la República de Corea se realizó en dos etapas. En la primera, se empalmaron las series del periodo comprendido entre 1990 y 1997. Para las series posteriores al año base, 1996 y 1997, los resultados finales de cada año fueron modificados mediante la aplicación de todos los cambios generados por el cambio de año base. Para las series de 1990 a 1994, se usó el método de la interpolación. La diferencia entre los resultados enlazados y los resultados originales compilados de 1995 fue interpolada a partir de la serie 1990-1994. En la segunda etapa, la serie de 1970 a 1989 fue extrapolada. Al extrapolar las series anteriores a 1990, el contenido y la cobertura del año base, junto con las tasas de crecimiento de la serie antigua, se trataron de mantener.

Constant price (1996=100)
estimate of component *i* for year *t* =

$$\frac{\text{Volume index of component } i \text{ for year } t}{\text{Volume index of component } i \text{ for year } 1996} \times \text{Constant price estimate of component } i \text{ for } 1996$$

Las estimaciones del PIB y sus componentes a precios constantes (1996=100) para los años anteriores a 1993 se revisaron independientemente de acuerdo con las respectivas tasas de crecimiento anual de la serie antigua. De esta manera las diferentes series a precios constantes conforman una serie continua con la estructura de precios del periodo final.

En Tailandia el método empleado puede resumirse en unos pocos pasos:

- Realizar un estudio no tan detallado, solo enfocándose en la producción a un nivel máxi-

La DSEC (The Macau Statistics and Census Services) ha adoptado las estimaciones a precios constantes base-fija como medida de volumen. Todas las estimaciones a precios constantes de las Cuentas Nacionales de Macao se obtienen por el método de deflación de precios. Actualmente, los precios de 1996 son usados para las estimaciones a precios constantes en las cuentas de gastos (el año base anterior era 1989). A fin de mantener actualizada la estructura de precios, el DSEC actualiza el año base periódicamente; el más reciente cambio de base se realizó de 1989 a 1996. En la estimación a precios constantes, las cifras del año 1993 y posteriores fueron recalculadas acorde con la estructura de precios relativos del nuevo año base. Para los años anteriores a 1993, en orden de mantener la continuidad, las estimaciones de los diferentes componentes del PIB son enlazadas por el método de empalme, empleando la siguiente fórmula:

mo de 11 sectores, ya que un estudio preliminar encontró que cuanto más detallada sea la investigación más difícilmente se puede llegar a realizar una conexión entre las series (debido a las variaciones en las clasificaciones y en la cobertura de detalles de cada elemento).

- Construir un Deflactor Implícito de Precios (IPD) para cada sector:

$$IPD = \frac{VA \text{ a precios corrientes } (t)}{VA \text{ a precios constantes } (t)} \times 100$$

- Conectar las series del IPD con las tasas de crecimiento, utilizando el método del indicador.

- Aplicar el método de la tasa de crecimiento al PIB a precios constantes base 1988 por sector.
- Hallar el PIB a precios corrientes multiplicando el resultado obtenido por todos los sectores de los hallados en cada IPD dividiendo por 100.
- Al contemplar que la serie obtenida es bastante larga (46 años), la estructura de la producción pudo haber cambiado en cierta medida, conduciendo a que la calidad del estudio no sea la mejor.

Sri Lanka expone su primer intento por empalmar el PIB a precios constantes para la serie 1975-1999 utilizando las bases 1975 y 1990. El método empleado fue la tasa de variación, a partir de los indicadores implícitos del PIB a precios constantes. Una vez calculado el indicador en ambos periodos, se realizó el encadenamiento de las series tomando como referencia la base 1990. Aducen que es el método más apropiado para empalmar las series debido a que la discrepancia se mantiene en torno al 5%. La serie empalmada del PIB fue validada como la serie control, en el momento que se presentaron mayores discrepancias en las estimaciones agregadas, donde se empalmó por separado cada una de las actividades del PIB.

Posteriormente procedieron a explicar las discrepancias, identificando los cambios metodológicos y posteriormente recalcularon los deflatores implícitos de precios en la serie empalmada, para garantizar su exactitud respecto a la serie original.

En Singapur el cambio de base se realiza cada 5 años. El empalme se lleva a cabo a partir de los índices encadenados; es decir, la aplicación del método de la tasa de variación, que preserva las tasas de crecimiento asociadas a las bases

anteriores. Sin embargo, al empalmar años lejanos al nuevo año base, no agregan el total debido a la pérdida de la propiedad de aditividad. Explican las diferencias en torno al impacto de los cambios de precios estructurales, las reclasificaciones y los cambios conceptuales que se presentan de una base a otra.

2.2. Métodos y alternativas empleadas por países subdesarrollados y en vías de desarrollo

Algunos de los países ubicados en esta categoría han realizado trabajos de encadenamiento que aportan importantes herramientas en la investigación, por tratarse de economías con características similares a la colombiana. A continuación se describen los hallazgos más relevantes.

El Instituto Nacional de Estadísticas y Censos de Argentina explica el procedimiento empleado en la retropolación del Valor Agregado Bruto (VAB) a precios básicos para el periodo comprendido entre 1980.I y 2006.IV. La técnica seleccionada en el caso de Argentina se denomina Método de la tasa de variación y consiste en aplicar las tasas de variación calculadas en la serie con base anterior, al nivel de la serie determinado por el nuevo año base. Ya que el nuevo Sistema de Cuentas Nacionales en Argentina con año base de referencia 2004⁶ valora la producción por sector de actividad a precios básicos (sin los impuestos a los productos); contrario a lo que sucede en la serie base 1986, donde la producción se expresaba a precios de mercado (incluyendo todos los impuestos) y en la base 1993, donde la valoración se realiza a precios de productor (es decir, sin incluir el

IVA facturado), surge la necesidad de eliminar de la base 1986 el total de los impuestos a los productos y del valor agregado bruto de la base 1993, los impuestos específicos y el impuesto a los ingresos brutos a fin de garantizar que las series empalmadas base 1986 y base 1993 se valoren a precios básicos.

Un ejercicio previo al trabajo de empalme de las series fue la reclasificación de las actividades económicas de la base 1986 (*con clasificación CIIU Rev2*), para hacerla comparable con la establecida en la base 1993 (*CIIU Rev3*). Tras realizar dicha homologación, se construyeron series agregadas en niveles e índices de volumen físico base 1993 para el periodo comprendido entre el primer trimestre de 1980 y el cuarto trimestre de 2006, a partir de las cuales se calcularon las tasas de variación interanuales (q) trimestrales para el periodo 1981.I-1993.IV.

Teniendo estas tasas, se procedió a aplicar el *método de la tasa de variación* a cada serie del valor agregado, aplicando la siguiente fórmula:

Donde:

VAB_{i-4}^{93} : Nivel del valor agregado bruto base 1993 del trimestre $i-4$.

VAB_i^{93} : Nivel de valor agregado bruto base 1993 del trimestre i .

q_v : Tasa de variación en tanto por uno entre trimestre $i-4$ y trimestre i del periodo 1981.I-1993.IV obtenida con el IVF base 1993 para 1980.I-2006.IV.

Para realizar el empalme de las series de SIFMI (servicios de intermediación financiera medidos indirectamente) y de Impuestos de los productos se emplearon las siguientes ecuaciones:

$$SIFMI_{i-4}^{93} = SIFMI_i^{93} / (1 + q_s)$$

$$Tp_{i-4}^{93} = Tp_i^{93} / (1 + q_{Tp})$$

Como resultado de este trabajo se obtuvo la serie trimestral 1980.I-2006.IV, base 1993 en CIIU Rev. 3, del Valor Agregado Bruto a precios básicos por sector de actividad económica.

Para este mismo país se realiza una recopilación de series históricas del producto y del ingreso por parte de la CEPAL. En este caso la variable de interés es el PIB para el periodo 1900-1997. Según el autor Ricardo Martínez, para construir una serie empalmada entre diferentes bases pueden seguirse varios criterios. El primero consiste en respetar los índices existentes en la serie antigua y aplicarlos a la nueva serie. La segunda supone que la diferencia se ha venido generando a lo largo de todo el periodo de manera acumulativa.

En este trabajo se utilizaron estos dos métodos en la construcción de la serie 1900-1997. El primer procedimiento se empleó para empalmar las bases 1950 y 1960; posteriormente con el segundo método se llevó a cabo el enlace de la base 1970 con la base 1986. En la primera etapa del ejercicio, comprendida entre los años 1900 y 1970, se aplicaron las variaciones de los índices existentes a dos dígitos de la clasificación CIIU, y el producto se obtuvo por agregación⁷. En el periodo restante, 1970-1997, se estimaron las variaciones entre los años 1971 y 1979 distribuyendo las desviaciones en forma acumulativa.

El Instituto Brasileño de Geografía y Estadística también se refiere al tema a través de la metodología empleada en la retropolación de la serie 1995-1999. En esta ocasión el estudio se extiende a las variables producción, consumo intermedio, consumo final y demás componentes del gasto. Concretamente este documento presenta el proce-

⁶ Los años base anteriores del PIB a precios constantes fueron: 1935, 1950, 1960 y 1970.

⁷ El enlace se realizó con la mayor desagregación posible, considerando que la evolución de una variable elemental se mantiene en términos interanuales con algún grado de homogeneidad.

dimiento adoptado para estimar la serie de las Tablas Recursos-Usos (TRU) de 1995 a 1999, valorada a precios corrientes y a precios constantes del año anterior, a partir de la TRU de 2000 a precios corrientes de la nueva serie del Sistema de Cuentas Nacionales (SCN).

El SCN de 1990 a 2003 fue publicado con sus TRU para 80 productos y 43 sectores de actividad (N80); mientras que la nueva serie está compuesta por 110 productos y 56 actividades. Para facilitar el encadenamiento de las dos series, se estableció que las TRU de la nueva serie tuvieran la misma clasificación de la serie anterior (a partir de 1995). Se supuso que las variaciones en volumen y precio de las operaciones de bienes y servicios de la clasificación N80, entre 1995 y 2000 se mantuvieron inalteradas, siendo modificadas solo cuando se trabaja en el reequilibrio oferta-demanda de las TRU⁸.

La tabla de producción del año 2000, a precios constantes de 1999, fue obtenida deflactando la tabla del año 2000 (de la nueva serie) a precios corrientes, por los índices de precio de la producción, margen de comercio, margen de transporte e impuestos por producto de la serie anterior⁹. La tabla de producción de 1999 a precios corrientes se obtuvo por retropolación de la tabla de producción de 2000 a precios de 1999, mediante los índices de volumen de la producción, margen de comercio, margen de transporte e impuestos por producto.

El consumo intermedio de las actividades y la demanda final a precios corrientes de 1999 fueron calculados a través de dos procedimientos diferentes. El consumo de las fami-

lias y la FBKF fueron deflactados por el índice de volumen por producto de las respectivas operaciones (dividir cada línea por el respectivo índice de volumen). El consumo intermedio fue el único caso en que se adoptaron índices de volumen por actividad. De esta manera, el cuadro de consumo intermedio por actividad a precios corrientes de 1999 fue obtenido dividiendo cada columna a precios constantes de 1999 por el respectivo índice de volumen de la actividad. Los componentes del valor agregado y las ocupaciones fueron divididos en tres segmentos: i) ocupaciones, ii) salarios e ingreso mixto y iii) contribuciones y otros impuestos y subsidios sobre la producción.

Las ocupaciones fueron obtenidas a través de la variación de la base actual del SCN y por la situación del empleo, de acuerdo con la actividad. Los salarios y los ingresos mixtos de 1999 se calcularon teniendo en cuenta las variaciones del valor de los salarios medios registrados en la base actual del SCN, aplicadas sobre la misma variable del año de referencia 2000. Con relación a las contribuciones y los impuestos y subsidios, los valores corrientes obtenidos en cada año fueron los registrados por la Administración Pública. El EBE se estimó por saldo, deduciéndose la remuneración y los impuestos netos del valor agregado resultante de las fases anteriores de la retropolación.

A partir de la TRU de 1999 a precios corrientes el proceso se repitió hasta estimar la TRU de 1995 a precios de 1994.

En una nota metodológica posterior, el Instituto de Estadística de Brasil examina el encadenamiento de la serie del VAB para los años comprendidos entre 1995 y 2001. Esta vez se menciona que la retropolación de una serie de cuentas a partir de un nuevo año base incorpora el cálculo de los agregados y

la nueva estructura de ponderaciones de las actividades económicas. Además de ello, siempre que sea posible es importante introducir los cambios conceptuales de la nueva serie en los años anteriores, procurando de esta manera aumentar la homogeneidad de las dos series. Sin embargo, por diversos motivos no es posible incorporar a los años anteriores todas las modificaciones introducidas en el año de referencia.

La estructura de las cuentas nacionales con año base de referencia 2000 se compone de 17 actividades económicas, una desagregación mayor en comparación al año de referencia 1985 que constaba de 15 actividades. Teniendo esto en cuenta, para realizar la retropolación de la serie con referencia 2002, se revisó la correspondencia entre las actividades de los dos años base (2002 y 1985). Como resultado se llegó a un total de nueve actividades económicas: Agricultura; Industria manufacturera; Producción y distribución de electricidad y gas, agua, aguas residuales y urbanas; Construcción civil; Comercio; Intermediación financiera, seguros y de pensiones y servicios relacionados; Administración, salud y educación pública; Seguridad social y Otros servicios.

La retropolación de las cuentas regionales del nuevo año de referencia fue estimada a partir del año 2002 retrocediendo hasta 1995, suponiendo que las variaciones en volumen y precio de las actividades en la clasificación propuesta (entre 1995 y 2002) se mantuvieron inalteradas. La estimación de los valores constantes para cada actividad (los valores a precios del año anterior) se llevó a cabo deflactando el valor agregado bruto corriente del año T, teniendo en cuenta el respectivo índice de precio para el año T-1 de la serie con referencia en 1985.

A su vez, el cálculo del valor corriente del año anterior (T-1) fue

⁸ Los cambios metodológicos implementados en la serie 1995-1999 implicaron la reestimación de algunos de los índices inicialmente adoptados.

⁹ Para el consumo intermedio y la demanda final a precios constantes del año anterior, se siguió el mismo procedimiento. Por su parte, la variación de existencias por producto no fue deflactada debido a la falta de sentido económico en la deflación de una variación.

retropolado aplicando al valor agregado bruto del año T a precios de T-1 (valor constante de T), su respectivo índice de volumen de la serie de referencia en 1985. Este procedimiento se llevó a cabo hasta el primer año de la serie regional retropolada (1995), luego fue ajustada a la serie retropolada de Cuentas Nacionales.

Chile es uno de los países que lidera la aplicación de técnicas estadísticas de retropolación. Uno de los documentos del Banco Central donde se hace un recuento de la metodología adoptada en el empalme de las series anuales y trimestrales del PIB para el período 1986-1995 base 1996 señala que al realizar la implementación del SCN 1993, la oficina de las cuentas nacionales realizó una revisión de los datos de series previas para obtener unos resultados consistentes frente al nuevo año base. Para ello, siguieron tanto recomendaciones del SCN 1993 como del SEC 1995, entre otras prácticas internacionales.

Después de una exhaustiva revisión de experiencias pasadas en retropolación propias, 1960-1973 base 1965 y posteriormente 1974-1980 base 1979 y otras tantas internacionales, se hizo una evaluación metodológica a partir de lo propuesto por Blades en su documento «Manteniendo las series consistentes de las Cuentas Nacionales» y Hexemberg en su artículo «Implementación del SCN93 revisión retrospectiva de los datos de las cuentas nacionales», lo cual fue determinante para que Chile optara por adoptar un método mixto de empalme de series combinando la *tasa de variación* con la *interpolación*, de modo tal que fuese un procedimiento simple que garantizara estimaciones reproducibles por los usuarios, tal como lo planteó el SEC.

Para las series anuales y trimestrales a *precios constantes* 1986-1995 se aplicó el método de la *tasa de variación*. De tal forma que «el nivel del trimestre i-4 base 1996 se obtiene dividiendo el nivel del trimestre "i" base 1996 por la tasa de variación entre los trimestres "i-4" e "i" en base 1986, respectivamente. Por ejemplo, el cuarto trimestre del año 1995, se calcula aplicando al cuarto trimestre del año 1996 base 1996, la tasa de variación entre el cuarto trimestre del año 1995 e igual trimestre del año 1996 en base 1986 y así sucesivamente hacia atrás» (Correa y otros 2002). Uno de los inconvenientes hallados al aplicar el método de la *tasa de variación* es el rompimiento del *principio de aditividad* que establece que la suma de los componentes del gasto debe ser equivalente al PIB.

En el caso de las series anuales y trimestrales a *precios corrientes* 1986-1995 se aplicó el método de interpolación entre lo registrado en cada uno de los años base (1986 y 1996) para un año común, definiéndolos como valores representativos por tratarse de mediciones efectuadas durante el cálculo de un año base.

Siendo V_{86}^t y V_{96}^t los valores nominales del año t en las bases 86 y 96 respectivamente, se hace necesario distribuir de manera gradual en el tiempo la siguiente diferencia de niveles (D) detectada para el año 1996:

$$D = \frac{V_{96}^{96}}{V_{86}^{96}}$$

Para distribuir esta diferencia de niveles durante el período 1986-1996 se calcula un promedio geométrico «r» donde:

$$r = D^{1/10}$$

Finalmente los valores nominales para calcular cualquier año t entre 1986 y 1996 se obtienen como:

$$V_{96}^t = V_{96}^t \times r^{t-1986}$$

Al realizar interpolaciones para el total del PIB y los componentes del gasto se produce una brecha entre ambos.

En el caso de las *series trimestrales a precios corrientes* para el PIB global y por componentes del gasto para el período 1990-1995¹⁰, el cálculo se realiza como el producto entre los registros a precios constantes (calculados anteriormente) y los deflatores trimestrales, calculados estos mediante el método de *tasa de variación*. Los resultados obtenidos para cada variable son luego ajustados proporcionalmente de modo que la suma de los valores trimestrales sea igual al valor anual obtenido mediante el procedimiento de interpolación. Finalmente, las series a precios corrientes trimestrales arrojan una diferencia entre la suma de los componentes y el total del PIB anual, por lo cual se ajustan al nivel anual.

La mezcla de estos dos métodos genera un ajuste en los deflatores, debido a que la serie a precios constantes se obtiene mediante extrapolación y no usando el método de deflación, mientras que la serie a precios corrientes se obtuvo por interpolación entre dos niveles correspondientes a los respectivos años base.

El documento advierte que: «Cuando a partir de ese empalme se pretenden efectuar *aperturas* en cualquier dirección (de transacciones, actividades, sectores o formas de variación), se imponen requerimientos de consistencia contable, sectorial e intersectorial a los datos. En este punto el ejercicio de empalme supera los límites del método aplicado y se transforma en un proceso de retropolación detallado con las dificultades prácticas que ello implica. Además, se debe considerar que un ejercicio de ese tipo

¹⁰ La serie de cuentas nacionales base 1986 no contiene estimaciones nominales trimestrales para el período anterior a 1990.

no garantiza la solución de incongruencias estadísticas que inevitablemente se producen al empalmar períodos alejados del año base definido» (Correa et al., 2003, p. 80).

Advierten también sobre la claridad que se debe tener ante los cambios conceptuales y estadísticos a los que la serie empalmada es susceptible al momento de contrastarse con la serie original. Los primeros se generan a partir del uso de nuevos conceptos, definiciones y clasificaciones; mientras que los estadísticos son producto de la utilización y adaptación de nuevas fuentes de información, la aplicación de otros métodos de empalme y procedimientos de estimación internos.

La pérdida de representatividad de esta serie empalmada para el caso chileno se reflejó básicamente en los siguientes aspectos:

- Variaciones significativas en los precios relativos y su estructura económica a medida que se iban alejando del año base.
- Altas tasas de inflación para los períodos 1960-1973 base 1977.
- La homologación y adaptación de la base más reciente a la más antigua:
 - Tratamiento de la administración pública.
 - El tratamiento dado a los servicios financieros imputados.
 - La diferencia de la valoración a la construcción.
 - El tratamiento de la tributación a los impuestos.

¹¹ Algunos países de la OCDE, posterior a ejecutar el SCN93, elaboraron reprocesos detallados para períodos recientes, no mayores a 5 años; mientras que para períodos anteriores aplicaron métodos de empalme estadístico.

Al concluir la revisión de su primera experiencia en retropolación, descubrieron que el reproceso no respeta las variaciones a precios constantes de la serie anterior para los períodos más cercanos a dicha base¹¹. Esto debido a omisiones generadas por las amplias necesidades de información y las dificultades presentadas en la definición del grado de representatividad de las estructuras de producción y sus precios relativos. Entre tanto, su experiencia con la serie 1974-1984 base 1986 definió como metodología global el uso de métodos estadísticos para empalmar el PIB y los componentes del gasto a precios corrientes y constantes. Allí realizaron una depuración del IVA contrastando la MIP77 con la MIP 86, así mismo fue necesario hacer una homologación de productos y actividades. Con la interpolación de variables en algunos casos remotos se realizaron algunos pequeños ajustes que permitiesen conservar la coherencia histórico-económica de las series.

Otra de las experiencias chilenas se refiere al empalme del PIB y los componentes del gasto empleando series anuales y trimestrales para el periodo 1986-2002 con año base de referencia 2003. La serie anual siguió los parámetros mencionados anteriormente; en el caso de las series de frecuencia trimestral el empalme se realizó mediante la aplicación de una variante del método de desagregación temporal propuesto por Denton (1971).

Para las series reales anuales el proceso de empalme aplicado implica recalcular las series a precios constantes y corrientes, junto con las series de deflatores. Por los mismos cambios en la escala de medición (precios) se hace necesario rebasar las mediciones ya realizadas en una base anterior, de tal manera que se nivelan las cuentas a la medición del año base más reciente, y retropolar la serie en

función de las tasas de variación implícitas en la antigua base de medición. En otras palabras, cambiar el nivel de las series, manteniendo sus tasas de variación anual:

Donde,

I: {0, ..., n-1}, siendo *n* el número de años de vigencia de la referencia previa.

R: Nivel de la variable en pesos constantes.

*b*₁: Base de referencia anterior.

*b*₂: Nueva base de referencia.

t: Año de referencia de la nueva base, *b*₂.

Con esta ecuación se realiza una translación proporcional de las series, de forma tal que las tasas porcentuales de variación se conservan homogéneas en ambas bases; solo se presentan cambios de nivel.

A su vez, es evidente que la progresiva obsolescencia del marco de referencia hace lo propio sobre las series nominales anuales, pues cuando se obtienen nuevas mediciones estas difieren de las obtenidas en la referencia anterior para el mismo año por cambios en la cobertura y sesgos de medición de precios y/o cantidades. Esta pérdida de precisión es cuantificable al comparar mediciones de la misma variable, para el mismo año en distintas bases (Stranger-Pozo 2007).

Para dar solución a esta problemática tomando en consideración dicha progresividad, asignan un diferencial observado al final de una base y al comienzo de otra de manera progresiva durante la vigencia de la medición:

$$N_{t-i}^{b_2} = N_{t-i}^{b_1} \times \left(N_{b_2}^{b_2} / N_{b_1}^{b_1} \right)^{(n-i)/n}$$

Donde,

I: {0, ..., n}, siendo *n* el número de años de vigencia de la referencia previa.

N: Nivel de la variable en pesos corrientes.

b₁: Base de referencia anterior.

b₂: Nueva base de referencia.

t: Año de referencia de la nueva base, b2.

La diferencia hallada entre las dos series es distribuida geoméricamente modificando los niveles y las tasas de variación.

Dentro de la metodología planteada se menciona el tema de los deflatores, los cuales son tratados como residual; es decir, después de obtenidos los resultados anteriores como series en valores nominales y reales «empalmadas» su razón corresponde a los nuevos deflatores.

$$D_{t-i}^b = \left(N_{t-i}^b / R_{t-i}^b \right)$$

Donde,

I: {0, ..., n}, siendo n el número de años de vigencia de la referencia previa.

N: Nivel de la variable en pesos corrientes.

R: Nivel de la variable en pesos constantes.

B: Base de empalme.

t: Año de referencia de la base de empalme, b.

«La limitación es la no aditividad de las series empalmadas. Es así que siguiendo el método de tasas de crecimiento se rompe la identidad contable entre el PIB y sus componentes, de manera que la suma de estos difiere del PIB empalmado en forma independiente. La recomendación del SCN 93 es sincerar dicha diferencia, denominada residuo, publicándola junto con los resultados del empalme. La interpretación de los resultados

obtenidos a través de este método es que el residuo es absorbido por variaciones de precios, o dicho de otro modo, que todo el ajuste por el cambio estructural asociado al nuevo año de referencia es asumido por los precios» (Stranger-Pozo 2007).

Chile también realiza una descripción del empalme de las series regionales del PIB para el periodo 1960-2001 con base 1996. Básicamente consiste en aplicar el método de la tasa de variación con el fin de mantener una consistencia metodológica con la actual serie histórica del PIB.

«Uno de los problemas que se presenta al aplicar este método, es que a medida que los años están más alejados del año base, mayores pueden ser las distorsiones y, por lo tanto, menor es la representación del resultado encontrado. Sin embargo, dada la situación de información existente en nuestro país, ninguna de las metodologías aplicables estará exenta de este tipo de errores» (Díaz, 2004). Al realizar una comparación entre el PIB regionalizado (suma de los PIB regionales) y el PIB encuentran algunas dificultades como la existencia de diferencias de ajuste, omisiones y discrepancias como consecuencia de la metodología empleada.

A pesar de que Ecuador no ha realizado un ejercicio juicioso de empalme de series, el Banco Central ha contemplado la posibilidad de realizar una retropolación de las series históricas de los cuadros de utilización, tanto a precios corrientes como constantes, para que a partir del empleo de índices de encadenamiento, se conserven los comportamientos económicos que se han venido registrando. Formalmente se pretende mantener las relaciones de equilibrio a fin de asegurar los cálculos del PIB a partir del gasto, la producción,

el ingreso, el cálculo del stock del capital y la elaboración de las cuentas trimestrales, entre otros. Se espera que los resultados de estos estudios de retropolación sean una herramienta importante en el diseño y establecimiento de políticas económicas.

México, en cambio, ha llevado a cabo en varias ocasiones el encadenamiento de sus series históricas del PIB. Un primer ejercicio se refiere al empalme de la serie 1970-2001. Al respecto se menciona que la primera serie de tiempo publicada por el INEGI abarcó el periodo 1970-1985, y tuvo como base la Matriz de Insumo-producto de 1970. Luego, se realizaron trabajos para integrar la matriz insumo-Producto de 1980, que serviría como base de la nueva serie de tiempo comprendida entre los años 1980-1995. En 1993 el INEGI realiza la actualización de los niveles del PIB, la estructura económica y las ponderaciones, dando como resultado la serie de tiempo actual que va desde 1988 a 2001 y cuya base de ponderación son los precios del año 1993.

Debido a las diferencias observadas en las series históricas retropoladas y extrapoladas del PIB, resulta necesario realizar un ejercicio sobre el encadenamiento de series, que arroje valores homogéneos y consistentes. En este caso de la serie del PIB total a precios corrientes, fue estimada mediante el método de interpolación entre los niveles alcanzados por las series en 1970, 1980 y 1993. Se considera que los valores extremos utilizados para interpolar son representativos por ser mediciones realizadas durante el cálculo de un año base. Las series son tomadas por segmentos según bases continuas, siendo V_{70}^t y V_{80}^t los valores nominales del año "t" en las bases 1970 y 1980, de manera que es necesario distribuir gradualmente en el

tiempo la diferencia de niveles (D) detectada para el año 1980, representada por:

$$D = \frac{V_{80}^{80}}{V_{70}^{80}}$$

Para distribuir esta diferencia de niveles durante el periodo 1970-1980, se calcula el promedio geométrico r como

$$r = \frac{\ln(V_{70}^{80}) - \ln(V_{70}^{70})}{t_{80} - t_{70}}$$

Así, se obtiene una serie de tendencia auxiliar entre 1970 y 1980, base 1970, expresada como:

$$A_{70}^t = V_{70}^t + r$$

Las fluctuaciones de la serie original se calculan como se señala a continuación:

$$d_{70}^t = V_{70}^t - A_{70}^t$$

Luego se estima la tendencia sustituida:

$$T_{80}^t = \frac{\ln(V_{80}^{80}) - \ln(V_{70}^{70})}{t_{80} - t_{70}}$$

Finalmente, los valores nominales para cualquier año t entre 1970 y 1980 se obtienen mediante la siguiente ecuación:

$$V_t = T_{80}^t + d_{70}^t$$

La serie histórica entre 1980 y 1993 se completa siguiendo el mismo procedimiento, con la diferencia que el año inicial en la base 1993 es el nivel alcanzado en 1988.

Por su parte, la serie del PIB total a precios constantes se calculó encadenando las tasas de variación real de las series originales. Los niveles fueron estimados mediante el método de la tasa de variación. De esta manera se comienza por segmen-

tar los periodos de las series base 1970 y 1980. El primer periodo inicia con la tasa de variación real de 1970 y se elige un año en particular, que no esté muy alejado del año base, pero que evidencie cambios relevantes de la estructura económica. El segundo segmento inicia con la tasa de variación real de 1978 (base 1980) y finaliza con la tasa de variación real del año 1988. Para completar esta última serie se toman las variaciones reales de la serie 1989-2001 base 1993.

A partir de la serie completa de las tasas de variación real desde 1970 al año 2001, los niveles del PIB total son calculados descontando la tasa de variación real del nivel del año base 1993.

Debido a que la serie a precios constantes se obtiene mediante retropolación y no por el método de deflación, mientras que la serie a precios corrientes se estima por interpolación entre los niveles correspondientes a los respectivos años base, se genera un ajuste en los deflatores. Una de las limitaciones de este ejercicio es que a medida que los cálculos se alejan del año base la diferencia entre el PIB publicado por el INEGI y el estimado por el encadenamiento, se hace proporcionalmente mayor.

El Instituto Nacional de Estadística y Geografía de México también se ha ocupado de retropolación series del PIB trimestral a precios corrientes. En principio se trata la serie 1993-2002. Los resultados del PIB trimestral del Sistema de Cuentas Nacionales de México (SCNM) utilizan el Sistema de Clasificación Industrial de América del Norte (SCIAN). La información estadística básica empleada tiene su origen en investigaciones como: las encuestas industriales, del comercio, de la construcción, de los servicios y del empleo, además de estadísticas sobre la minería no petrolera, las asociadas al transporte, y los

registros administrativos de empresas públicas y privadas. Igualmente, se utilizaron índices de precios obtenidos de diversas fuentes, tales como el índice nacional de precios de productor, mayoristas, consumidor, construcción, salario mínimo general y otros índices generados a partir de diversas fuentes.

Bajo el principio de que en un marco contable no es posible conservar las relaciones entre un agregado y sus componentes, si ambos se empalman por separado, porque se pierde el principio de aditividad, fue necesario trabajar con todas las series y obtener los agregados por suma de los valores parciales. De esta manera, se utilizó un método que pasó por la selección de 737 indicadores de precios obtenidos de diversas fuentes. Los índices de precios utilizados para expandir trimestralmente el producto real son tipo Laspeyres, con ponderación fija en el año base. Es importante anotar que las estimaciones del producto trimestral pueden presentar un leve desvío respecto a los resultados finales que se obtienen en las mediciones anuales del SCNM.

El ejercicio de retropolación planteado, que consiste en la construcción de series de valores y precios para cada una de las 737 clases de actividad, solo proporciona coherencia temporal analítica y parcial a las variables seleccionadas. Es decir, el ejercicio solo es posible para el PIB trimestral a precios corrientes, sin que se pueda extender a sus componentes. De esta manera, si a partir de este empalme se pretenden realizar aperturas (de transacciones, actividades, sectores o formas de variación), se imponen requerimientos de coherencia contable sectorial e intersectorial a los datos.

Dando continuidad a esta serie, a continuación se realizan los trabajos necesarios para conciliar la serie trimestral 1993-2002. En este

ejercicio se llevó a cabo la construcción de series para cada una de las 737 clases de actividad SCIAN a partir de la información estadística básica disponible que permitió dar el tratamiento de reproceso detallado a 199 clases, aplicar métodos de proporción e interpolación entre años de referencia para 408 clases y, finalmente, el uso del método del indicador para 130 clases¹².

Aunque la literatura de México es extensa, no es muy profunda en lo que se refiere al tratamiento dado a las series antes de aplicar los métodos de retropolación mencionados. No es muy evidente en los estudios realizados la inclusión de los cambios metodológicos ocurridos entre las bases anteriores y el año de referencia actual.

Países como Nicaragua han realizado ejercicios de empalme con variables como la construcción. El método aplicado a las series con frecuencia trimestral fue la tasa de variación, teniendo en cuenta las áreas efectivamente construidas en Managua y el resto de departamentos. De esta manera, el nivel del trimestre $i-4$ de la serie empalmada se obtuvo dividiendo el nivel del trimestre i , que recoge la nueva información, entre la tasa de variación comprendida entre los trimestres $i-4$ e i de la serie existente de áreas en cabeceras departamentales. Por ejemplo, el segundo trimestre del año 2004 para Managua se obtuvo aplicando al segundo trimestre del año 2005 de la serie nueva, la tasa de variación entre el segundo trimestre de 2005 y el segundo trimestre del año 2004 de la serie existente, y así sucesivamente hacia atrás. Una vez aplicado el método de la tasa de variación, se obtuvo el nivel de las áreas trimestrales efectivamente

te construidas. Posteriormente se calculó la estructura porcentual de los destinos económicos en cada zona geográfica de la serie existente. El procedimiento anterior arrojó como resultado áreas de construcción trimestrales con una cobertura ampliada, en las cuales se respetó tanto la evolución de las series existentes, como la distribución a lo interno de cada zona geográfica.

Paraguay continúa en la misma línea de investigación y apunta hacia el empalme de las series del PIB de 1982 a 1994 desde el punto de vista del gasto, tanto en guaraníes constantes como corrientes con base 1982 y la nueva base 1994. El procedimiento empleado se denomina interpolación de datos, ya que se utiliza información histórica de la serie anterior para obtener los datos de la nueva serie hacia atrás. Se dispone de información para el periodo 1991-2003 sobre la serie de datos del PIB por el lado del gasto, con la nueva base 1994. También se cuenta con el PIB total en guaraníes constantes de 1962 a 2004, el cual fue obtenido mediante la sumatoria de los componentes retropolados del VAB de cada rama de actividad económica.

Para preservar la coherencia de la información, se procedió a realizar el empalme de los datos a partir de las participaciones en el producto. Así, en el periodo comprendido entre 1991 y 2003 se realizaron los cálculos de las participaciones de cada uno de los componentes por el lado del gasto, tanto en guaraníes constantes como corrientes, para las dos series de distinto año base. Una vez calculadas todas las participaciones del consumo, inversión, exportaciones e importaciones, respecto al PIB, se procedió a comparar estos ratios para las dos bases. Esto se realiza para observar si la media es relativamente estable en el periodo considerado. De ser así, se emplea la media ob-

tenida para reproducir el ratio de la nueva base hacia atrás.

Se realizó el cálculo de los deflatores implícitos del producto para los dos años base y, finalmente, para mantener la consistencia con los datos agregados del PIB tanto en guaraníes corrientes como constantes de la nueva base, se emplearon las importaciones totales como variable residual.

Uruguay es uno de los países que sobresale en esta investigación por el contenido metodológico de sus trabajos de retropolación de series de cuentas nacionales. En la nota realizada, Jorge Ponce describe el procedimiento empleado para empalmar la serie 1955-2000 con frecuencia anual y trimestral para el PIB y los componentes del gasto. Se parte de una revisión de la información histórica de cuentas nacionales y, a fin de mantener la consistencia en las series de tiempo, evalúan las alternativas metodológicas planteadas por Hexeberg (2000); el reproceso detallado, ya que considera todas las posibles fuentes de inconsistencia al compilar nuevamente las cuentas¹³ y por supuesto las técnicas de empalme estadístico.

A pesar de que el reproceso detallado constituye la opción más precisa, se encuentra que su adopción es poco factible debido a los requerimientos de información y recursos que impone esta técnica. Un error de estimación a la hora del recálculo de las series podría estar ampliando las diferencias presentadas por el empalme estadístico, que tampoco está exento de problemas, ya que como se mencionó anteriormente algunas propiedades contables se rompen cuando se empalman los componentes y el agregado por separado.

En el desarrollo de este documento, se realiza un empalme de series a precios constantes utilizando el mé-

¹² El marco conceptual de este documento es el mismo que aparece señalado en el caso de la retropolación a precios corrientes.

¹³ En tanto, inconsistencias debidas a cambios en clasificaciones o valoraciones no son atendidas por las técnicas de empalme.

todo de la tasa de variación¹⁴ con el fin de construir una serie larga y comparable que mantenga las características temporales de las originales. Para las series trimestrales se detectó que la información de corto plazo (estacionalidad y efecto calendario) contenida en las series base se conservaba. En cuanto a las series anuales, la metodología consiste en estimar los valores empalmados para $t = k + h - 1$ con:

$$C_{k+h}^{t=k+h-1} = C_{k+h}^{t=k+h} \times (C_k^{t=k+h-1} / C_k^{t=k+h})$$

Y así secuencialmente.

Donde:

k y k+h: Años base

C_k^t y C_{k+h}^t : Valores para el momento t en pesos constantes de las bases k y $k+h$

Este procedimiento determina la pérdida de la propiedad de aditividad, generándose una diferencia entre el PIB empalmado y la suma algebraica de los componentes del gasto empalmados por separado. En esta forma, se detecta una diferencia media de 0,19%, con una diferencia máxima hallada en el año 1973 de 2,91%.

Por otra parte, el empalme de las series a precios corrientes es realizado con el método de interpolación. Partiendo de tener en consideración que las cuentas nacionales se compilan con el mayor grado de detalle y confianza en los años base, los datos de los años 1961, 1983 y 1988¹⁵ pueden ser considerados robustos y, por tanto, fueron tomados como referencia¹⁶. Una vez establecidos estos puntos, el problema del empalme consiste en distribuir gra-

dualmente la diferencia que para el nuevo año base se origina entre los valores calculados con la base anterior y la nueva. Siguiendo las recomendaciones de Hexeberg (2000) y Correa *et Al.* (2002 y 2003):

$$D = V_{k+h}^{t=k+h} / V_k^{t=k+h}$$

Donde,

k y k+h: Dos años base consecutivos

V_k^t y V_{k+h}^t : Valores a precios corrientes en el momento t de la serie objetivo en las bases k y $k+h$

Para distribuir esta diferencia se aplica el promedio geométrico:

$$r = D^{1/10}$$

Seguidamente se hace el cálculo del valor empalmado para el año t en la nueva base mediante la siguiente expresión:

$$V_{k+h}^t = V_k^t \times r^{(t-k)}$$

Tras obtener los resultados del empalme para el PIB de la serie 1955-2002, una vez más las propiedades contables inherentes a la consistencia de las cuentas nacionales son pasadas por alto debido al empalme. En este caso la media de las diferencias porcentuales alcanza -0,06% y la diferencia mayor es de -0,70 en el año 1980.

En el documento también aplican una metodología alternativa que consiste en ajustar un modelo de regresión a la serie en la base más reciente contra su homóloga en la base inmediatamente anterior, para luego estimar la serie empal-

mada utilizando los coeficientes del modelo¹⁷. Ya que no se garantiza la obtención de la consistencia contable y transversal, y tras solo haber alcanzado cierta consistencia temporal analítica, plantean algunas conciliaciones para las series empalmadas a efectos de dotarlas de tales consistencias.

La conciliación transversal de los datos se realiza con el fin de devolver la propiedad de la aditividad a las series empalmadas. Para conseguirlo se aplica una versión modificada de minimización cuadrática propuesta por Denton (1971).

Venezuela inicialmente muestra las dificultades que pueden presentarse a la hora de realizar un empalme, entre ellas la disponibilidad de series oficiales del BCV que corresponden a años bases diferentes elaborados con diversas versiones del Sistema de Cuentas Nacionales de las Naciones Unidas¹⁸. Esto hace que las series no sean comparables entre sí. Francisco Rodríguez, autor del documento, hace referencia a las variaciones entre las series disponibles del PIB y a su procedencia:

- Hay diferencia en los datos primarios utilizados, dado que algunos investigadores utilizan datos diferentes a los oficiales.
- Diferencias en los años bases utilizados, particularmente derivadas del mayor peso que los precios relativos del año X le dan a las variaciones del PIB.
- Las diferencias originadas por la técnica de empalme estadístico adoptada.

¹⁴ Empalme de las series anuales a precios constantes del PIB y los componentes del gasto, así como del IVF PIB trimestral.

¹⁵ Si bien no se realizó un cambio de base, la revisión de las series B83 para el año 1988 y siguientes (b8388) representa un incremento en la confianza de las estimaciones en ellas contenidas.

¹⁶ Estableciendo a estos años como las playitas del cálculo o puntos de referencia para realizar la interpolación.

¹⁷ Tal metodología presenta el inconveniente, para las series anuales, de contar con insuficientes observaciones, ya que los periodos por los cuales se compilan ambas series suelen ser cortos (en general inferiores a los diez años). Por otro lado, y si bien la insuficiencia de datos suele ser un problema de menor entidad para el ajuste de series de mayor frecuencia, se opta por mantener la homogeneidad y simplicidad que el método de la tasa de variación provee para el empalme de tales series.

¹⁸ La serie base 1957 fue elaborada de acuerdo con Naciones Unidas (1960), mientras que las series posteriores han sido elaboradas de acuerdo con United Nations (1968). Actualmente, el BCV se encuentra preparando un nuevo conjunto de series macroeconómicas con año base 1997, utilizando la última versión del Sistema de Cuentas Nacionales, que data de 1993.

La solución planteada para los dos últimos consiste en la presentación de un nuevo indicador del PIB real para Venezuela (índice encadenado de Fischer), en el cual las variaciones anuales son ponderadas por la participación en el PIB de cada sector en los años en los cuales ocurre la variación, además incluye ajustes por variaciones en la canasta de productos y por reclasificación de productos en el PIB.

La práctica internacional ha mostrado una inclinación hacia la utilización de índices encadenados¹⁹ para el cálculo de sus principales agregados de cuentas nacionales. Para este ejercicio la información utilizada provino del BCV, con la cual se busca hallar el índice encadenado de Laspeyres:

$$X_t^j = \prod_{i=1}^j \frac{\sum_{i=1}^N p_i^t x_i^{t+1}}{\sum_{i=1}^N p_i^t x_i^t} = \frac{\sum_{i=1}^N p_i^t x_i^{t+1}}{\sum_{i=1}^N p_i^t x_i^t} = \prod_{i=1}^j \sum_{i=1}^N n_{it} Y_{it}$$

Así como también el índice encadenado de Törnqvist:

$$X_T^j = \prod_{t=1}^N \prod_{i=1}^N (Y_{it})^{1/2(n_{it} + n_{u+1})}$$

Siendo:

n_{it} : Participación del sector i en el PIB nominal del periodo t .

γ_{it} : Tasa de crecimiento del PIB en el sector i a precios constantes en el periodo t .

Los sectores utilizados serán las actividades económicas que forman la base de la subdivisión productiva del producto interno bruto: petróleo crudo y gas natural, refinación de petróleo, agricultura, minería, industria manufacturera, electricidad y agua, construcción, transporte,

almacenamiento y comunicaciones, comercio, restaurantes y hoteles, productores de servicios del gobierno general, y resto de la economía. Adicionalmente, forman parte del PIB los derechos de importación, servicios imputados a las instituciones financieras y ajuste por unificación del tipo de cambio, los cuales serán tratados como sectores económicos adicionales para efectos del cálculo.

Estos sectores, añade el autor, corresponden a los bienes i en la implementación empírica del índice encadenado de Laspeyres y el de Törnqvist, por lo que el índice presentado en el documento es solo una aproximación al que se lograría de disponer con información precisa de n_{it} y de γ_{it} para cada bien agregado en el PIB.

Los inconvenientes hallados al encontrar un cambio de base son saltos discretos que ocurren en el PIB nominal en el año en que ocurre el cambio; es decir, el porcentaje de participación de cada sector puede variar según la base que se observe debido al surgimiento y la desaparición de algunos bienes y servicios, y su reclasificación. Por otra parte, estas diferencias de nivel también son aducidas a cambios metodológicos entre una base y otra.

Ante la falta de información sobre el crecimiento de la fracción de la producción no observada en cada sector antes del cambio de año base, el autor asume que había crecido a la misma tasa que la fracción observada. A partir de ello se recomienda aplicar al PIB a precios corrientes un método de empalme simple cuando los cambios estructurales carecen de relevancia²⁰, en el cual la tasa de crecimiento $\gamma_{it} = Y_{it} / Y_{it-1}$ del sector i en el año t es:

$$\gamma_{it} = \frac{x_{it-1}}{x_{it}}$$

Donde:

γ_{it} : PIB empalmado para el sector i en el año j .

X_{it} : PIB a precios constantes del sector i en el año t .

Para el caso, donde la variación originada por el cambio de base del PIB a precios corrientes se deba a cambios estructurales, entonces se recomienda distribuir proporcionalmente estos cambios entre los años durante los cuales ocurrió el surgimiento y/o desaparición de algunos sectores. En ese caso la tasa de crecimiento $\gamma_{it} = Y_{it} / Y_{it-1}$ del sector i en el año t será:

$$\gamma_{it} = \frac{x_{it}}{x_{it-1}} (n_{ib0})^{(1/b-s)} \quad 8$$

Donde:

$\eta_{ibs} = PIB_{ib}^b / PIB_{is}^b$ denota la proporción del PIB a precios corrientes con el nuevo año base y con el viejo año base.

PIB_{ix}^b : es el PIB nominal correspondiente al sector i en el año b con año base x .

Finalmente, este fue el empalme geométrico utilizado en este documento, debido a que los cambios metodológicos comúnmente no son más que reflejos de cambios estructurales. La mayoría de los cambios metodológicos introducidos por el BCV en su cambio de base a 1984 tienen que ver con la mejor cobertura de la información correspondiente al sector público, la mejora en los cálculos de establecimientos con menos de cinco personas, y la aplicación de índices específicos para estimar por separado las ventas externas, internas y VE de la industria petrolera.

Por último, se mencionarán algunos países asiáticos que han llevado a cabo ejercicio de empalme mediante la aplicación de métodos de deflación o extrapolación. Entre estos se encuentra Bangladesh,

¹⁹ El movimiento hacia el uso de índices encadenados ha cobrado particular relevancia a partir del auge reciente en los sectores ligados a la tecnología de información en los países desarrollados, debido a que índices basados en ponderaciones obtenidas en un año base específico distorsionaban la contribución de estos sectores debido a las fuertes variaciones en precios que habían presentado en un espacio reducido de tiempo (Daveri, 2002).

²⁰ Los cambios en el PIB a precios corrientes que ocurren con los cambios de año base reflejarían simplemente una mejor capacidad de la nueva metodología para capturar la producción existente.

que dispone de tres series de cuentas nacionales con años base 1972-73, 1984-85 y la más reciente versión base 1995-96. El ejercicio de enlace de las bases se ha hecho siguiendo un enfoque mixto; para los sectores primarios se ha recurrido a la revaluación, para los sectores secundarios la estimación de los precios corrientes se ha llevado a cabo utilizando deflatores de precios apropiados y para los servicios, se ha recurrido a la deflación o la extrapolación del volumen dependiendo de la disponibilidad de indicadores físicos y deflatores de precios apropiados. En cuanto a los sectores manufactureros, Bangladesh ha intentado realizar cinco ejercicios utilizando enfoques y conjuntos de datos alternativos relacionados con el censo de las industrias manufactureras (CMI), el índice anual de la producción industrial (QIP) y el índice de precios al productor (IPP). En uno de los ejercicios el método de encadenamiento también fue probado; esto es, la relación entre el índice de volumen para determinado año sobre el índice de volumen del año base, por la estimación a precios constantes del año base.

$$\frac{\text{Volume index of component } i \text{ for year } t}{\text{Volume index of component } i \text{ for year } 1995 - 96} \times \text{constant price estimate of component } i \text{ for } 1995 - 96$$

De cierta forma Bangladesh ha preferido el método de revaluación, la deflación de precios y la extrapolación del volumen en diferentes sectores, en lugar del empalme. Esto se ha traducido en un cambio de las tasas de crecimiento en la nueva serie enlazada. Igualmente para realizar el enlace hacia atrás de las categorías del gasto del PIB, Bangladesh ha seguido el criterio de deflación de precios haciendo uso de los deflatores adecuados.

Indonesia cuenta con cuatro series del ingreso nacional para los siguientes años base: 1960 es el año de referencia para la serie del PIB 1960-1973; 1973 corresponde

al año base para el periodo 1973-1983; el año base 1983 para la serie del PIB 1983-1993, y finalmente el año base 1993 para la serie del PIB 1993-2000. Las dos últimas series del PIB han incorporado mejoras en el alcance y la cobertura de las actividades gracias a la utilización de las tablas I-O (Insumo-Producto). Las series enlazadas se encuentran disponibles para el periodo 1978-1993 con año base de referencia 1983.

El método de enlace consiste en la extrapolación de los índices de volumen y técnicas de empalme. Para el nuevo año base 1993, no se ha llevado a cabo algún ejercicio de empalme. Actualmente se encuentra en el proceso de compilar índices encadenados anuales.

En Nepal se considera la posibilidad de cambiar la base para la estimación a precios constantes de 1994/95 y enlazar las series para todo el periodo. Durante el ejercicio no ha sido relevante mejorar las estimaciones del año de referencia 1994/95; en cambio, se ha trabajado en mantener los niveles de las tasas de crecimiento en la última serie como se encontraban en la serie más antigua. Además, se han realizado esfuerzos por dividir los principales sectores en actividades o grupos de productos por la disponibilidad de datos en las nuevas estimaciones. Así, hay 27 sub-actividades/grupos de productos en los nueve sectores. La pro-

ducción y el CI para cada subsector han sido identificados a precios corrientes y se han deflactado para obtener las estimaciones a precios constantes. Se espera fomentar la construcción de índices de precios, realizar la asignación del consumo de capital y contar con las estimaciones de los impuestos a nivel de sector o subsector.

A diferencia de los países asiáticos mencionados, Vietnam realizó el empalme de sus series con el método del indicador, teniendo en cuenta una mejor cobertura, la identificación de nuevos contenidos de medición y el método de valoración de los indicadores anuales. Inicialmente su sistema de Cuentas Nacionales era regido por el MPS (Material Product System) y no por el UNSNA (United Nations System of National Accounts), los cuales difieren principalmente por su interpretación del valor agregado y de la acumulación de stocks de riqueza. Esta situación genera diferentes resultados a la hora de extrapolar los agregados. En el MPS muchos servicios son considerados como no generadores de valor agregado y, por lo tanto, son excluidos de la producción total. EL MPS solo mide la producción anual de bienes materiales, en contraste con los servicios.

En 1991, se hizo el cambio de MPS a SNA, lo cual condujo a un nuevo método en el cálculo de las cuentas como:

$$PIBa \text{ precios constantes} = \text{Producción total} - \text{Consumo intermedio}$$

$$PIBde \text{ 1993 a precios de 1994} = \frac{PIB \text{ 1994 a precios corrientes}}{K}$$

$$K = \frac{PIB \text{ de 1994 a precios de 1989}}{PIB \text{ de 1993 a precios de 1989}}$$

$$PIBde \text{ 1992 a precios de 1994} = \frac{PIB \text{ 1993 a precios de 1994}}{H}$$

$$H = \frac{PIB \text{ de 1993 a precios de 1989}}{PIB \text{ de 1992 a precios de 1989}}$$

En Pakistán los avances en temas de empalme no han sido muy evidentes, el año base fue encadenado para las series 1964-65, 1969-70 y

1975-76. En estas cuentas solo se realizó la vinculación de cambios metodológicos a cada serie y no se aplicó en ningún momento un pro-

cedimiento de empalme diferente al encadenamiento de series: cambio de año base de 1980-81 a 1999-2000 en junio de 2002 y Cuentas Nacionales base 1999-2000 encadenada a la base 1980-81.

3. Conclusiones y alcances de la investigación

Si bien la actualización de las bases estadísticas provee un redimensionamiento de la realidad económica, la literatura consultada pone en evidencia la importancia de implementar métodos estadísticos que permitan empalmar los diferentes años de referencia, procurando en la medida de lo posible incorporar los principales cambios metodológicos implementados en la base más reciente. Al respecto puede mencionarse que los cambios más evidentes se concentran en la valoración de la producción, la medición de los servicios de intermediación financiera (SIFMI) y la reclasificación de las actividades económicas.

Las metodologías internacionales revisadas dejan claro que este es un tema novedoso y de constante investigación para varios países que por ahora se encuentran en un proceso de adaptación al SCN o que, por el contrario, se limitan a contemplar la posibilidad de considerarlo en el tratamiento de sus series estadísticas. Otros, en cambio, lideran la aplicación de los métodos de empalme, como Uruguay, Chile, México, Brasil, Argentina, España, y otros países europeos que en su mayoría aplican modelos econométricos complejos.

El método de empalme más utilizado es la interpolación en el caso de las series a precios corrientes, que consiste básicamente en partir del año base actual y distribuir las diferencias generadas a lo largo

de la serie mediante un promedio geométrico, manteniendo inalterados los niveles de los años de referencia. Para las series a precios constantes se recomienda hacer uso del método de la tasa de variación, en el que se intenta aplicar las tasas de crecimiento de la serie anterior a los niveles obtenidos en el año base actual.

Un aspecto que sobresale en los documentos estudiados es la importancia de incorporar los cambios metodológicos actuales en las series anteriores para garantizar resultados coherentes con la realidad económica, sin alterar el curso de la historia. Al respecto, el problema más frecuente que suele presentarse es la falta de información, ya que aunque no se trate del método de reproceso detallado en el que se persigue básicamente rehacer las cuentas nacionales, resulta elemen-

tal disponer de series históricas que faciliten la aplicación de las técnicas mencionadas. El método del indicador se propone como una alternativa para superar este inconveniente en la medida en que puede recurrirse a información relacionada con la variable de interés que facilite el empalme de los datos.

Un aspecto en común entre los países consultados consiste en las variables empleadas para los ejercicios de empalme. En su mayoría suelen orientarse a calcular grandes agregados macroeconómicos como el PIB, el Valor Agregado, el consumo intermedio, la producción y algunos componentes de la demanda. Esto debido a que trabajar a niveles tan detallados puede imponer requerimientos estadísticos que exceden los alcances de los métodos mencionados.

Tabla 1. Métodos de empalme usualmente empleados internacionalmente.

Métodos de enlace de series históricas			
	Tasa de variación	Interpolación	Econometría
Países	Argentina		
	México	México	
	Brasil	Paraguay	
	España	España	España
	Francia		
	Chile	Chile	
	Canadá		
	Ecuador		
			Europa
	Nicaragua		
	Uruguay	Uruguay	
	Venezuela	Venezuela	
Variables		PIB	
		VAB-VA	
		producción, CI, CF, componentes del VA,	
		componentes del gasto,	
		empleo y población.	

Fuente: Elaboración propia.

Aunque no es considerada propiamente una limitación, debe mencionarse que las técnicas de interpolación y tasa de variación generan una pérdida de aditividad de las series empalmadas, ya que se rom-

pe la identidad contable del PIB por sumatoria de sus componentes. Sin embargo, este no es un obstáculo, ya que, como sucede actualmente en la presentación de las Cuentas Nacionales de Colombia a precios

constantes por encadenamiento, dicha discrepancia puede evidenciarse en los resultados de manera que sea el usuario final quien determine a su juicio cómo distribuirla.

De acuerdo con los resultados de la investigación realizada, la propuesta planteada para llevar cabo los trabajos de retropolación de las Cuentas Nacionales de Colombia consiste en emplear una técnica mixta de empalme para el PIB y los componentes del gasto aplicando el método de interpolación, para las series anuales a precios corrientes, y el método de la tasa de variación, para las series anuales a precios constantes. Para esto se sugiere, aparte de realizar revisión exhaustiva de la estadística e información disponible, incorporar los cambios metodológicos relevantes que se han presentado en los diferentes años base, pues es claro que en agricultura, minería, industria y servicios han aparecido nue-

vos conceptos, fuentes y formas de cálculo que no pueden omitirse en el empalme de las series. Sumado a esto, debe haber una conciliación entre las diferentes bases respecto a la clasificación utilizada y la manera de valorar la producción, ya que, por ejemplo, mientras en la base 1975 se registra un total de 35 ramas de actividad y la producción se valora a precios productor, en la nueva base las actividades económicas ascienden a 61 y la producción se expresa en términos de precios básicos.

Finalmente, a partir de la investigación y los hallazgos encontrados en cada caso, el Grupo de Retropolación de la Dirección de Síntesis y Cuentas Nacionales del DANE publicó recientemente los resultados del empalme de la serie 1975-2010 para los principales agregados macroeconómicos y grandes ramas de actividad de las Cuentas Nacionales de Colombia

base 2005, atendiendo los requerimientos de los usuarios en cuanto a la necesidad de disponer de una serie de tiempo amplia y comparable que refleje de la mejor manera los hechos más relevantes de la historia económica nacional. La retropolación de la serie se realizó conciliando los resultados históricos de las bases 1975, 1994 y 2000 con los más recientes avances en temas de conceptos, cambios metodológicos y de clasificación incorporados en el año de referencia actual. En particular, fueron cuatro los cambios metodológicos incorporados, a saber: cambios en la valoración de la producción, inclusión de los SIFMI en las bases precedentes y la medición de los cultivos ilícitos. Los resultados del PIB obtenidos son coherentes con las distintas bases, tanto en niveles como en tendencia, garantizando de esta manera la oportunidad del método empleado en los trabajos de retropolación.

Bibliografía

Ajmal, Muhammad (13-16 de febrero de 2001). Rebasing and Linking of National Accounts Series February 13-16, 2001 at Bangkok, Thailand. Federal Bureau of Statistics, Islamabad, Pakistán.

Banco Central de Chile, Departamento de Cuentas Nacionales (s. f.). Interpolación del PIB: Un ejercicio para datos chilenos [online]. Recuperado el 2 de marzo de 2012 de http://www.eclac.cl/deype/noticias/noticias/5/33615/2008_07_taller-MetodoEstadisticos_BCh2.pdf.

Banco Central de Nicaragua-Departamento de Cuentas Nacionales Trimestrales. Nicaragua. Metodología de Empalme de área

efectivamente construida [online]. Recuperado el 1 de marzo de 2012 de http://www.bcn.gob.ni/publicaciones/eventuales/metodologias/metodologia_empalme.pdf.

Blades, Derek (marzo de 2000). Maintaing consistent time-series of National Accounts" Joint ADB/ ESCAP Workshop on Rebasing and Linking of National Accounts Series, 21-24, Bangkok, Thailand.

Bureau Economic Analysis-U.S. Department of Commerce (s. f.). GDP and Other Major NIPA Series, 1929-2011 [online]. Recuperado el 24 de febrero de 2012 de http://www.bea.gov/scb/pdf/2011/08%20August/0811_gdp_nipas.pdf.

Centro de Estudios de las Finanzas Públicas. (2003). Encadenamiento de series históricas del PIB de México 1970-2001. Cámara de Diputados, CEFP/012/2003. Ciudad de México.

Chong, Frederick (13 a 16 de febrero de 2001). Macao National Accounts Statistics. Macao Statistics and Census Services, Macao, China.

Colmán, Humberto (octubre de 2005). Empalme de las series del PIB por el lado del gasto. Subsecretaría de Estado de Economía e Integración-Dirección de Estudios Económicos núm. 4/2005. Paraguay.

Comisión de las Comunidades Europeas, Fondo Monetario Interna-

- cional, Organización para la Cooperación y Desarrollo Económicos, Naciones Unidas y el Banco Mundial (1993). Sistema de Cuentas Nacionales 1993. Departamento de Información Económica y Social y Análisis de Política-División de Estadística, Bruselas/Luxemburgo, Nueva York, París, Washington D. C.
- Comisión de las Comunidades Europeas, Fondo Monetario Internacional, Organización para la Cooperación y Desarrollo Económicos, Naciones Unidas y el Banco Mundial (2008). Sistema de Cuentas Nacionales 2008. Departamento de Información Económica y Social y Análisis de Política-División de Estadística, Bruselas/Luxemburgo, Nueva York, París, Washington D. C.
- Contas Nacionais (15 de octubre de 2007). Retropolação da base 2000 das Contas Nacionais Anuais. INE-Destaque, Lisboa.
- Correa, Escandón, Luengo y Venegas, José (abril de 2003). Empalme de series anuales y trimestrales del PIB. Banco Central de Chile. *Revista Economía Chilena*, 6(1). Chile.
- Correa, Escandón, Luengo y Venegas, José (septiembre de 2002). Empalme de series anuales y trimestrales del PIB. Banco Central de Chile, Working Papers Central Bank of Chile, Documento de Trabajo núm. 179 Chile.
- Davery, Francesco. (2002): "The new economy in Europe, 1992-2001", Oxford Review of Economic Policy, vol. 18, nº 3.
- De la Fuente, Ángel (diciembre 2009). A Mixed Splicing Procedure for Economic Time Series. CESifo Working Paper núm. 2876. España.
- De la Fuente, Ángel (septiembre de 2009). Series Enlazadas de Algunos Agregados Económicos Nacionales y Regionales, 1955-2007. Versión 2.1. España.
- Denton, Frank (marzo de 1971). Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization. *Journal of the American Association*, 66(333), 99-102. Canadá.
- Department of Census & Statistics (13 a 16 de febrero de 2001). Work programme in the implementation of 1993 SNA. Ministry of Finance and Planning, Colombo, Sri Lanka.
- Di Fonzo, Tommaso (2003). Constrained reinterpolation of high-frequency data using related series. European Commission, 2003 Edition, Luxembourg.
- Díaz, Javier (diciembre de 2004). Empalme Series de PIB Regionales 1960-2001, Base 1996. Ministerio de Planificación y Cooperación-División Planificación Regional. Chile.
- Dirección de Estadística Económica (abril de 2011). Justificación, fuentes de información, sectorización institucional, nomenclaturas de productos e industrias, clasificaciones de transacciones de bienes y servicios, de distribución, capital, financieras, otros flujos, balances y saldos contables adoptados para el cambio de año base 2007. Banco Central de Ecuador, Cuadernos de Trabajo núm. 133. Ecuador.
- Dirección Nacional de Cuentas Nacionales (julio de 2001). Resumen Metodológico para la Medición del PBI por Departamentos. INEI, Lima.
- Economic and Social Commission for Asia and the Pacific, (13 a 16 de febrero de 2001). Rebased and Linking of the GDP Series from the base year 1984/85 to 1994/95. Nepal.
- Feu, Aumara y Feu Carlos, ECE. (s. f.). **Retropolação das Contas Nacionais até 1947 [online]**. Recuperado el 7 de marzo de 2012 de http://ecen.com/eee64/eee64p/re-trapolando_contas_nacionais.htm.
- Hexeberg, Barbro (mayo de 2000). Implementación del SCN 1993: Revisión retrospectiva de los Datos de las Cuentas Nacionales. NSD, SNA News, N°11.
- Hwan, Kang (13 a 16 de febrero de 2001). National Accounts Compilation in the Republic of Korea. The Bank of Korea, Seoul, Republic of Korea.
- INEGI (2011). Metodología Sistema de Cuentas Nacionales de México. Instituto Nacional de Estadística y Geografía. *SCNM*, x, 186 p. il. México D. F.
- INEI (julio de 2000). Metodología para el Cambio de Año Base de las Cuentas Nacionales del Perú. *Metodologías Estadísticas*, año 1, núm. 9. Lima.
- Instituto Brasileiro de Geografia e Estadística (2002). *Revisão da Série 1995-1999 (retropolação)*. IGBE-CONAC, Nota Metodológica núm. 22. Brasil.
- Instituto Brasileiro de Geografia e Estadística (2002). *Revisão da Série 1995-2001 (retropolação)*. IGBE-CONAC, Nota Metodológica núm. 26. Brasil.
- Instituto Nacional de Estadística y Censos (INDEC). (s. f.). Nota metodológica [online]. Recuperado el 14 de febrero de 2012 de http://www.indec.gov.ar/nuevaweb/cuadros/17/pbi_80-05_metodologia.doc.
- Instituto Nacional de Estadística y Geografía. México D. F. (s. f.). Historia del Sistema de Cuentas Nacionales de México (1938-2000) [online]. Recuperado el 28 de febrero de 2012 http://unstats.un.org/unsd/wsd/docs/Mexico_wsd_HistoriaSCNM.pdf.
- Instituto Nacional de Estadística. (s. f.). Nota Metodológica: Contabilidad Regional de España, base 2000. Serie homogénea 1995-2006 [on-

- line]. Recuperado el 29 de febrero de 2012 de http://www.ine.es/daco/daco42/cre/meto_homog.pdf.
- Kolli, Ramesh y Sinhg, Gulab (13 a 16 de febrero de 2001). Status os implementation of 1993 SNA in India and Rebasng and Linkng of Indian National Accounts Statistics. Central Statistical Organisation, New Delhi, India.
- Lee, Pete (noviembre de 2011). UK National Accounts – A Short Guide. Office for National Statistics. United Kingdom.
- Leung, Tammy (13 a 16 de febrero de 2001). Development of National Accounts Statistics in Hong Kong, China. Census and Statistics Department. Hong Kong, China.
- Lihua, Dong (13 a 16 de febrero de 2001). The Status of Implementation of the 1993 SNA in China. National Bureau of Statistics. Beijing, China.
- Lopes, Gabriela y Soares, Paulo (junio de 2004). Quarterly Series for the Portuguese Economy: 1977-2003. Banco de Portugal, *Economic Bulletin*. Lisbon.
- Maluquer, Jordi (2009). Del cosmos al caos: Una nueva serie enlazada del PIB de España entre 1950 y 2000. Universidad Autónoma de Barcelona, *Revista de Economía Aplicada*, 49 (XVII), 5-45. Barcelona.
- Marconi, Salvador y Salcedo, Jaime (junio de 1995). La acumulación de capital fijo en el Ecuador 1965-1993. Banco Central de Ecuador, Nota Técnica núm. 15. Ecuador.
- Martínez, Ricardo (abril de 1999). Recopilación de series históricas del producto y del ingreso. CEPAL LC/BUE/R.242, Buenos Aires.
- Naciones Unidas (1960): A system of national accounts and supporting tables, Nueva York, Naciones Unidas.
- Naciones Unidas (1968): System of National Accounts, 1968, New York, United Nations.
- OECD (s. f.). Methodology for gap filling [online]. Recuperado el 14 de marzo de 2012 de <http://www.eea.europa.eu/data-and-maps/indicators/gdp-outlook-from-oecd#methodology>.
- Office for National Statistics U.K. (s. f.). Historic National Accounts [online]. Recuperado el 16 de febrero de 2012 de urlm.in/ludz.
- Office of National Economic and Social Development Board (NES-DB). (13-16 de febrero de 2001). Description on the Rebasng Process Gross Domestic Product of Thailand: 1951-1996. NESDB, Bangkok, Thailand.
- Ponce, Jorge (noviembre de 2004). Una Nota sobre Empalme de Series de Cuentas Nacionales. Banco Central de Uruguay, *Revista de Economía, Segunda Época*, XI(2). Montevideo.
- Prados, Leandro (septiembre de 2009). Del cosmos al caos: La serie del PIB de Maluquer de Montes, Retropolación para el período 1993-2002, Base 2003. Retropolación para el período 1993-2002 Año Base 2003, Instituto Nacional de Estadística y Geografía-SCNM, México D. F.
- Prados, Leandro, (agosto de 2006). Assessing the Bias in Spliced GDP Series: Evidence from Spain's National Accounts, 1954-2000. Universidad Carlos III de Madrid. Madrid.
- Rodríguez, Francisco (julio a diciembre de 2004). Un nuevo índice encadenado del PIB de Venezuela, 1957-2001. *Revista BCV*, XVIII(2), 99-118. Caracas.
- Saleh, Kusmadi y Mangiri, Komet (13 a 16 de febrero de 2001). Compilation of National Account of Indonesia. BPS-Statistics Indonesia. Jakarta, Indonesia.
- SCNM (2009). Series del PIB Trimestral a Precios Constantes. Retropolación para el período 1993-2002 Año Base 2003. Instituto Nacional de Estadística y Geografía-SCNM, México D. F.
- SCNM (2010). Series del PIB Trimestral a Precios Corrientes. Retropolación para el período 1993-2002 Año Base 2003. Instituto Nacional de Estadística y Geografía-SCNM, 383 p. il. México D. F.
- SCNM (s. f.) Series Mensuales del Indicador Global de la Actividad Económica (IGAE).
- Secretaria de Estado de Gestão e Planejamento (SEGPLAN). (s. f.). *Retropolación [online]*. Recuperado el 21 de febrero de 2012 de <http://www.seplan.go.gov.br/sepim/pub/pcb/resultadoretopoladabr.pdf>.
- Singapore Department of Statistics (13 a 16 de febrero de 2001). Rebasng and Linkng of National Accounts Series. Singapore Department of Statistics, Singapore.
- Stranger, Michael (marzo de 2007). Empalme del PIB y de los Componentes del Gasto: Series Anuales y Trimestrales 1986-2002, Base 2003. Banco Central de Chile, *Estudios Económicos Estadísticos*, núm. 55, Chile.
- Universidad Carlos III de Madrid (junio de 2006). *Working Paper* 09-04. Madrid.
- Van, Nguyen (13 a 16 de febrero de 2001). Some issues in estimation of some integrated indicators and compilation main Accounts in Viet Nam. The General Statistical Office of Viet Nam, Viet Nam.
- Virola, Rómulo y Domingo, Estrella (13 a 16 de febrero de 2001). Chai-

ning the Philippine National Accounts Series, National Statistical Coordination Board (NSCB), Manila, Philippines.

Zoque, Md. Zobdul y Abdul, Mohammad (13 a 16 febrero 2001). Reba-

sing and Linking of National Accounts Series in Bangladesh. Bangladesh Bureau of Statistics, Bangkok.



**Cuenta de
emisiones al
aire: estudio de
caso emisión
de metano por
la fermentación
entérica del
ganado bovino**

Luz Dary Yepes Rubiano

Fuente: www.sxc.hu

Cuenta de emisiones al aire: estudio de caso emisión de metano por la fermentación entérica del ganado bovino

Luz Dary Yepes Rubiano: Ingeniera Forestal de la Universidad Distrital. Especialista en manejo de medio ambiente de la Universidad de Los Andes. Asesora de Cambio Climático del Grupo de Indicadores y Cuentas Ambientales de la Dirección de Síntesis y Cuentas Nacionales del DANE. Correo electrónico: ldyepesr@dane.gov.co

Fecha de recepción: 24 de enero de 2013
Fecha de aceptación: 16 de septiembre de 2013

Resumen: este documento de investigación tiene como finalidad proporcionar a los usuarios elementos conceptuales para entender y dimensionar el desarrollo de la contabilidad ambiental económica en términos de la cuenta de emisiones al aire, a partir de un ejercicio piloto de cuantificación de la emisión de gases de efecto invernadero (GEI) ilustrada por el cálculo de la emisión de metano por la fermentación entérica del ganado bovino para el periodo 2000-2010. Igualmente, se pretende indicar la necesidad de estructurar la cuenta de emisiones al aire como una herramienta a ser usada dentro de la implementación de la política de cambio climático para orientar la Estrategia Colombiana de Desarrollo Bajo en Carbono, y fortalecer los Inventarios Nacionales de Gases de Efecto Invernadero.

Palabras claves: emisiones al aire, fermentación entérica, gases de efecto invernadero, metano, cambio climático.

Abstract: the finality of this investigation document is to provide to the users some conceptual elements to understand and dimension the development of the SEEA in terms of the emissions to the air, starting from a pilot exercise of quantification of the emission of greenhouse gases (GHG) illustrated by the calculation of the methane emission produced by the enteric fermentation of the cattle for the period 2000-2010. Likewise, to indicate the necessity of structuring the Air Emissions Accounts, like a tool to be used in the implementation of the climate change policy for orienting the low Carbon Development Strategy and strengthening the National GHG Inventories.

Key words: air emissions, greenhouse gas, enteric fermentation, methane, climate change.

1. Introducción

La cuenta de emisiones al aire se estructura en el marco de la contabilidad ambiental económica. Esta cuenta permite dimensionar los impactos al medio ambiente en términos de emisiones de gases contaminantes derivados de las actividades económicas de producción, consumo y procesos de acumulación. La información económica que permite cuantificar estas emisiones proviene de los datos sectoriales que conforman las Cuentas Nacionales de Colombia (CN).

Para cuantificar las emisiones derivadas de las actividades económicas se deben surtir los siguientes pasos: i) establecer en unidades físicas las diferentes variables que determinan la actividad económica; ii) identificar los gases que se generan por cada actividad económica, clasificándolos de acuerdo con su impacto en el medio ambiente, como son: gases de efecto invernadero; acidificación, calidad del aire, precursores de la capa de ozono, metales pesados, compuestos orgánicos persistentes, entre otros; y iii) asociar a cada tipo de gas un factor de emisión.

Las emisiones se registran a través de la matriz de oferta-utilización de emisiones que permite relacionar por actividad económica de CN el tipo de gas emitido en unidades físicas, y que al ser cruzados con el valor de la producción de cada actividad en términos monetarios (matriz híbrida) se obtiene el indicador de intensidad de las emisiones, que permite establecer por cada unidad de valor producida cuántas emisiones se generan. Este indicador es usado generalmente en el tema de desarrollo sostenible para analizar las implicaciones ambientales de los patrones de producción y consumo de un país.

La cuenta de emisiones, dentro de los nuevos avances en el contexto de la contabilidad ambiental económica, se concibe para generar información que permita la toma de decisiones en torno a los efectos positivos y negativos que ejercen las actividades económicas sobre el medio ambiente; así, dentro del componente de cambio climático es de interés establecer las emisiones de gases efecto invernadero derivadas de las actividades económicas.

En el documento Conpes 3700, en el ítem «V. Lineamientos»-«Información para la toma de decisiones», se establece que «El diseño, formulación, seguimiento, evaluación económica y ajuste de la Estrategia Institucional para la Articulación de Políticas y Acciones en Materia de Cambio Climático en Colombia y sus componentes, deberá soportarse en información cuantitativa y cualitativa precisa, resultado de: i) la generación de escenarios climatológicos futuros a escalas nacional, regional y local, la cuantificación de los impactos físicos sobre ecosistemas y los servicios que estos suministran, sociedades y actividades económicas, y de la estimación de los costos y beneficios económicos y sociales de los mismos, y ii) la producción de información estadística básica requerida en los análisis de cambio climático y los resultados de los avances de las cuentas de emisión en el marco de la contabilidad ambiental».

Dentro del desarrollo de la propuesta conceptual de la cuenta de emisiones, se realiza un ejercicio preliminar para estimar la emisión de metano por la fermentación entérica del ganado bovino. Este ejemplo permite dimensionar los pasos a seguir para la cuantificación de los diferentes gases derivados por las actividades económicas, como son: la identificación de las necesidades de información a nivel de CN, la disponibilidad de

factores de emisión que reflejen las condiciones propias de las actividades económicas del país, la incertidumbre asociada tanto a la actividad como al factor de emisión utilizado que permite evaluar la calidad de los resultados obtenidos; realizar análisis comparativos en el ámbito internacional en términos de emisiones por actividad económica; y finalmente contabilizar las emisiones y su relación con el valor de la producción económica de un país o región.

Este documento se estructura en cinco secciones. En la primera se presentan los aspectos conceptuales claves para el desarrollo de la cuenta de emisiones; en la segunda sección se establece la metodología para el cálculo de las emisiones del ejercicio piloto; en la tercera sección se muestran los resultados de la estimación de las emisiones de metano por la fermentación entérica del ganado bovino en el ámbito de la generación de GEI; en la cuarta parte se presenta la relación de la información derivada de la cuenta de emisiones con la toma de decisiones, y por último, se encuentran las conclusiones.

2. Aspectos conceptuales¹

El Sistema de Cuentas Nacionales 2008 precisa que «las cuentas ambientales tienen por objeto describir, en un marco basado en el SCN, la repercusión de la utilización (y en ocasiones el agotamiento) de los recursos naturales y la generación de residuos que contaminan el aire y las aguas. Asimismo, permiten identificar las actividades específicas que se dedican a prevenir o combatir los efectos de la actividad humana sobre el medio ambiente».

La contabilidad ambiental económica debe considerarse como una cuenta satélite del SCN, en la cual

¹ Adaptado a partir de los documentos: i) System of Environmental Economic Accounts (SEEA) y ii) SNA 2008.

se distinguen cuatro diferentes tipos de flujos:

- Los **productos**: Son bienes y servicios producidos dentro de la esfera económica y que se utilizan dentro de esta, incluyendo el flujo de bienes y servicios entre la economía nacional y el resto del mundo.
- Los **recursos naturales**: Comprenden recursos minerales y energéticos, el suelo, el agua y los recursos biológicos.
- Los **insumos del ecosistema**: Comprenden el aire y los gases necesarios para la combustión y el agua como sustento de vida.
- Los **residuos**: Son los productos involuntarios e indeseados de la economía, cuyo precio es cero para quien los produce y pueden ser reciclados o descargados al medio ambiente. «Residuos» es el único término que se utiliza para hacer referencia a los desechos sólidos, efluentes (descargas al agua) y emisiones (descargas al aire).

La cuenta de emisiones al aire² es parte del flujo de residuos y tiene como objetivo establecer los flujos netos de residuos gaseosos y material particulado al aire, originado por la economía tanto por actividades de producción como de consumo y procesos de acumulación. Esta cuenta agrega las emisiones al aire directas procedentes de las ramas de actividad económica y del consumo final de los hogares; así como las emisiones al aire derivadas de los desechos sólidos entregados a gestores de residuos y/o depositados en instalaciones de tratamiento; y de los residuos lí-

quidos³ resultado de la producción y el consumo, y que se vierten directamente al medio ambiente y/o se entregan al alcantarillado para su tratamiento o disposición final.

2.1. Los gases emitidos

Las sustancias generadas por las ramas de actividad económica y el consumo final de los hogares, que se han identificado para cuantificar dentro de la cuenta de emisiones al aire de acuerdo a su impacto en el medio ambiente, son las siguientes (ver glosario):

- **Gases de efecto invernadero (GEI)**: Los GEI son gases que se generan principalmente por la combustión de los combustibles fósiles usados por las ramas de actividad económica y el consumo final de los hogares⁴, los procesos industriales que transforman materias por medios físicos o químicos, el uso de productos, el tratamiento de residuos (incineración, vertimiento, otros), la quema de los residuos agrícolas y cremación, la fermentación entérica, manejo del estiércol, aplicación de pesticidas y cal a los suelos y cultivos agrícolas. Los gases que se generan son dióxido de carbono (CO₂), metano (CH₄), óxido nitroso (N₂O), hidrofluorocarbonados, perfluorocarbonos y hexafluoruro de azufre, entre otros.
- **Acidificación (lluvia ácida)**: Los principales precursores de la lluvia ácida son los óxidos de azufre y los óxidos de nitrógeno, compuestos químicos generados a través de dos vías principales: las fuentes antrópicas, principalmente a partir de la quema de combustibles fósiles, o las

fuentes naturales, compuestas por emisiones volcánicas y fuentes termales, entre otras.

- **Calidad del aire**: Los contaminantes relacionados con la calidad del aire son el monóxido de carbono (CO), óxidos de azufre (SO_x), óxidos de nitrógeno (NO_x), ozono (O₃), plomo (Pb) y material particulado, los cuales se generan por la quema incompleta de los combustibles, combustión en fuentes fijas (calderas y hornos), el uso de compuestos de petróleo y solventes orgánicos en procesos industriales y de limpieza, quemas agrícolas y de origen doméstico, entre otros.
- **Precusores de ozono**: Los gases que hacen parte de esta categoría son los clorofluorocarbonos e hidrofluorocarbonos, sustancias utilizadas como refrigerantes, agentes espumantes (en fabricación de espumas de poliuretano y poliestireno principalmente), propelentes para aerosoles y solventes, como agentes de extinción de fuego, para protección de equipo electrónico, como reactivo químico en laboratorios, entre otros.
- **Metales pesados**: Los metales pesados se encuentran en forma natural en la corteza terrestre. Estos se pueden convertir en contaminantes si su distribución en el ambiente se altera mediante actividades humanas. En general esto puede ocurrir durante la extracción minera, el refinamiento de productos mineros o por la liberación al ambiente de efluentes industriales, y por la disposición de residuos metálicos. Son considerados metales pesados elementos como arsénico, cadmio, plomo, zinc,

² Los residuos que se devuelven al sistema económico no son contabilizados en la cuenta, por considerar que no impactan el medio ambiente.

³ La cantidad de los residuos líquidos derivados de las actividades económicas y de los hogares hacen parte de la contabilidad de la cuenta del agua.

⁴ Se incluyen las emisiones relacionadas con el uso de combustible para vehículos particulares y motocicletas, para la calefacción, cocción y agua caliente, y otras derivadas del uso de solventes utilizados para pintar, aerosoles por uso de sprays, por actividades de ocio como las quemas por uso de fogatas, quema de basuras en los jardines; cortadoras de césped, máquinas para cortar setos y otras maquinarias de jardinería (Eurostat, 2009).

mercurio, cadmio, cobre, berilio, flúor, níquel, entre otros.

- **Compuestos Orgánicos Persistentes (COP):** Los COP son compuestos orgánicos fabricados artificialmente, y de acuerdo con el tratado de Estocolmo los más peligrosos son los pesticidas (aldrina, DDT, clordano, hexaclorobenceno, entre otros), herbicidas, insecticidas organoclorados, los productos industriales como el bifenilo policlorado (PCB) y los productos secundarios no inten-

cionales de procesos industriales o combustión como son las dioxinas y los furanos.

2.2. Tabla oferta-utilización

Para el reporte de las emisiones se estructura la tabla de oferta-utilización, donde la oferta está dada por la generación de las emisiones por rama de actividad económica y el consumo final de los hogares, por tipo de sustancia; y la utilización representa las emisiones que salen directamente al medio ambiente, las emisiones que son capturadas

por medio de tecnologías para devolverlas al sistema económico (ejemplo: gas metano para la generación de electricidad) y el total del uso de las emisiones (tabla 1).

2.3. Los indicadores

En el desarrollo de la cuenta de emisiones al aire se derivan entre otros indicadores:

- Total de emisiones por tipo de gas por impacto por rama de actividad económica y consumo final de los hogares.

Tabla 1. Oferta-utilización de la cuenta de emisiones.

Tabla oferta-utilización cuenta de emisiones al aire	Unidades de masa: toneladas									
	Tabla de oferta				Tabla de utilización					
	Ramas de actividad económica				Consumo final hogares	Emisiones desde rellenos sanitarios (Acumulación)	Total de oferta de las emisiones	Flujos al medio ambiente	Emisiones liberadas al medio ambiente	Total uso de las emisiones
Tipo de sustancia	Agricultura	Minería	Industria manufacturera	...	Transporte	Otros				
Dióxido de carbono										
Metano										
Óxido nítrico										
Hidrofluorocarbonados										
Monóxido de carbono										
Metales pesados										
Material particulado										
Otros...										

Fuente: Elaboración propia a partir de la propuesta «table 3.6.1 Air emissions account» del documento System of Environmental-Economic Accounts (SEEA), 2012.

- Variación de las emisiones por tipo de gas por impacto por rama de actividad económica y consumo final de los hogares.
- Intensidad de las emisiones por tipo de gas por impacto por rama de actividad económica.
- Emisiones por unidad de valor agregado, según ramas de actividad.
- Emisiones por unidad del PIB.

3. Material y métodos-ejercicio piloto

Para cuantificar las emisiones derivadas de las 61 ramas de actividad económica de CN, la acumulación y el consumo final de los hogares se debe: i) establecer en unidades físicas las diferentes variables que determinan la actividad económica; ii) identificar los gases que generan estas actividades, clasificándolos de acuerdo con su impacto en el medio ambiente; y iii) cuantificar las emisio-

nes asociando a cada tipo de gas un factor de emisión (FE).

La metodología sencilla para estimar las emisiones consiste en combinar la información en unidades físicas de las ramas de actividad económica y el consumo final de los hogares con los coeficientes que cuantifican las emisiones. Por consiguiente, la ecuación básica es:

Emisiones = Información en unidades físicas X Factor de emisión (FE)

Por la actividad pecuaria se produce metano, producto de la fermentación entérica, como consecuencia del proceso digestivo de los herbívoros. Tanto los rumiantes (ganados vacuno y ovino), como algunos no rumiantes (cerdos y caballos), producen metano. La cantidad emitida está en función del sistema digestivo y del tipo de alimento consumido por el animal. Igualmente hay emisiones de metano y óxido nítrico derivados de la gestión del estiércol que obedece a la descomposición en condiciones anaeróbicas y depende de las técnicas utilizadas para su manejo.

El metano (CH₄) es producido por animales herbívoros, como un producto de la fermentación entérica, en un proceso digestivo por el cual los carbohidratos son digeridos por microorganismos en moléculas sencillas para ser absorbidas por el torrente sanguíneo. Tanto los animales rumiantes como los bovinos y los caprinos, como los no rumiantes como caballos y cerdos producen CH₄, aunque los rumiantes son la fuente más importante debido a la capacidad para digerir celulosa, por medio de microorganismos en el tracto digestivo. La cantidad de CH₄ liberado depende del tipo de ganado, edad y peso del animal; calidad y cantidad de comida, y energía gastada por el animal.

Para una mejor ilustración, en el ganado bovino se emite CH₄ debido a que el proceso digestivo ocurre bajo condiciones anaeróbicas en presencia de diferentes tipos de bacterias. Estas degradan la celulosa ingerida a glucosa, que fermentan después a ácido acético y reducen a dióxido de carbono, formando CH₄ en el proceso. De hecho, la emisión de CH₄ representa una pérdida de rendimiento en la actividad ganadera, porque la energía alimenticia se pierde en

forma de gas y no es aprovechada como leche o carne¹.

Las características de la dieta tienen un gran efecto en la producción de CH₄, de ahí que países con pocas limitaciones de alimentación para sus ganados reportan datos de menores emisiones de CH₄ y mayores eficiencias energéticas. Cuando los bovinos son alimentados con forrajes de baja calidad nutritiva, se presentan deficiencias en nutrientes esenciales para los microorganismos ruminales, por lo que la eficiencia en el crecimiento de estos en el rumen es baja. Bajo esta circunstancia, la producción de CH₄ podría representar entre el 15 y el 18% de la energía digestible. La corrección de estas deficiencias nutricionales podría reducir las pérdidas a valores cercanos al 7% (Bernal, 2007: 22).

3.1. Ecuación para calcular las emisiones de metano por la fermentación entérica del ganado bovino

Para ilustrar la metodología de cálculo de las emisiones en el contexto de la cuenta, a continuación se muestra un primer ejercicio para la estimación de las emisiones de metano por la fermentación entérica del ganado bovino en el ámbito de la generación de GEI:

$$\text{Emisiones (Gg CH}_4 \text{ año}^{-1}) = (N_{(\text{Gbovino})} * FE_{(\text{Gbovino})}) / 10^6$$

Donde,

Emisiones (Gg CH₄ año⁻¹): «El metano se produce en los herbívoros como subproducto de la fermentación entérica, un proceso digestivo, por el cual los microorganismos descomponen los carbohidratos en moléculas simples para la absorción en el flujo sanguíneo. La cantidad de metano que se libera depende del tracto digestivo, la edad

y el peso del animal, así como de la calidad y la cantidad del alimento consumido. Los rumiantes (p. ej., vacunos, ovinos) son fuentes importantes de metano». (Panel Inter gubernamental de Expertos sobre Cambio Climático, IPCC, 2006).

Las emisiones se presentan en términos de gigagramos (Gg) de metano por año. Los Gg⁵ son la unidad utilizada para el reporte de las emisiones de GEI.

(N_(Gbovino)): Cantidad de cabezas de ganado bovino por año.

El cálculo piloto de las emisiones de metano por el ganado bovino se realizó para el periodo 2000-2010p. Por lo tanto, se establece la cantidad de animales anuales (de ahora en adelante denominado *stock* de ganado), registrados en las Cuentas Nacionales base 2005.

El *stock* de ganado bovino se presenta con corte al 31 de diciembre de cada año, donde se incluyen los aumentos y disminuciones del *stock* a lo largo del año.

Para la estimación de las emisiones, el *stock* de ganado se divide en dos categorías: ganado de leche; y otro ganado que agrupa vacunos para carne, doble propósito y machos reproductores.

A continuación, en la tabla 2 se muestran los datos del *stock* de ganado bovino de Cuentas Nacionales, para cada año del periodo 2000-2010p.

p: provisional

FE_(Gbovino): factor de emisión para la población de ganado bovino, kgCH₄/cabeza/año. En el país por ahora no existen datos oficiales de factores de emisión de metano por el ganado bovino presente en el territorio. Por consiguiente, para el cálculo piloto se utiliza el factor de

¹ Gigagramo equivale a 1 kilotonelada (kt).

emisión para fermentación entérica de la región de Latinoamérica, propuesto en las directrices del IPCC

1996 para la elaboración de los inventarios nacionales de gases de efecto invernadero (tabla 3).

calcule la proporción entre la producción del ganado bovino y su respectivo valor agregado. En las Cuentas Nacionales este producto hace parte de la actividad 03 (Animales vivos, productos animales y productos de la caza), donde no se tiene específicamente el valor agregado de cada uno de sus componentes.

Tabla 2. Stock de ganado bovino. 2000-2010p.

Año	Tipo de ganado		Total stock	Variación porcentual
	Leche	Otro ganado		
2000	763.600	22.737.168	23.500.768	
2001	775.247	23.169.752	23.944.999	1,89
2002	781.422	23.870.055	24.651.477	2,95
2003	799.825	24.546.673	25.346.498	2,82
2004	792.919	24.921.910	25.714.829	1,45
2005	833.750	25.363.368	26.197.118	1,88
2006	826.841	25.761.194	26.588.035	1,49
2007	845.596	26.252.200	27.097.796	1,92
2008	852.495	26.881.020	27.733.515	2,35
2009	865.734	27.517.745	28.383.479	2,34
2010p	876.426	28.035.723	28.912.149	1,86

Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales-Cuentas Nacionales base 2005.

Tabla 3. Factores de emisión por fermentación entérica del nivel 1 para ganado vacuno.

Características regionales	Categoría de ganado	Factor de emisión	Comentarios
		(kg CH ₄ /cabeza/año) IPCC 1996	
América Latina: sector lechero comercializado en base a pastoreo. Hato de vacuno por separado en pasturas y tierras de pastoreo. Pequeña cantidad de alimentación en corral con granos. El ganado no lechero para engorde constituye una gran parte de la población.	Lechero: Emisiones de metano de vacunos que producen leche para el intercambio comercial y de terneros y vaquillonas que se crían para la producción láctea.	57	Producción promedio de leche de 800 kg/cabeza/año
	Otros vacunos: Emisiones de metano de todo ganado vacuno no usado para producción láctea, incluido: ganado vacuno cuidado o criado para la producción cárnica, animales de tiro y animales para la reproducción.	49	Incluye vacas, toros y animales jóvenes multipropósito.

Fuente: IPCC 2006: volumen 4, pág. 10-30.

3.2. Estimación de un indicador de intensidad de emisiones

La cuenta de emisiones en el marco de la contabilidad ambiental económica se considera como una herramienta para el análisis del impacto de las actividades económicas en el ambiente, en términos de las emisiones al aire. Uno de los indicadores clave para determinar el im-

pacto es el denominado «indicador de intensidad de las emisiones», el cual relaciona las emisiones de un gas con el valor agregado de la actividad económica.

Por ejemplo, para estructurar el indicador de intensidad de emisiones respecto al valor agregado del ganado bovino, se debería establecer el estudio pertinente en donde se

Por lo anterior el indicador, en teoría, estaría conformado de la siguiente manera: Indicador de intensidad de emisiones de gases de efecto invernadero respecto al PIB del ganado bovino:

IIEGEIGBi: GEI (ganado bovino) i/PIBGi

Donde:

- IIEGEIGBi: Indicador de intensidad de emisiones de gases de efecto invernadero respecto al PIB del ganado bovino en el periodo i (CO₂ equivalentes/\$).
- GEI (Ganado Bovino): Emisiones de gases de efecto invernadero del ganado bovino en el periodo i (CO₂ equivalentes).
- PIBGBi: Producto Interno Bruto del ganado bovino en el periodo i (\$).

3.3. Comparación de resultados

Los resultados de la emisión de metano por fermentación entérica del ganado bovino de este ejercicio piloto se comparan con respecto a cálculos semejantes: i) el reporte de las emisiones del país en el marco de la elaboración de su inventario nacional de GEI, y ii) el reporte de los países Anexo I y No Anexo I a la CMNUCC.

3.4. La evaluación del nivel de incertidumbre

En este aspecto es conveniente comentar el grado de incertidumbre de los datos fuente como son los facto-

res de emisión y los datos del stock. La incertidumbre en el ámbito de las emisiones de gases de efecto invernadero, de acuerdo con el IPCC, tiene la siguiente definición:

«Incertidumbre: es la falta de conocimiento del valor verdadero de una variable que puede describirse como una función de densidad de probabilidad (FDP) que caracteriza el rango y la probabilidad de los valores posibles. La incertidumbre depende del nivel de conocimiento del analista, el cual, a su vez, depende de la calidad y la cantidad de datos aplicables, así como del conocimiento de los procesos subyacentes y de los métodos de inferencia» (IPCC, 2006, vol. 1, cap. 3).

Estadísticamente definen la incertidumbre como «un parámetro asociado al resultado de la medición que caracteriza la dispersión de los valores que podrían razonablemente atribuirse a la cantidad medida, por ejemplo, la varianza de la muestra o el coeficiente de variación». Se explica entonces la incertidumbre como la imposibilidad de conocer con certeza absoluta el valor verdadero de las emisiones, las estimaciones pueden estar afectadas por factores no muestrales como errores en la medición, precisión de los instrumentos, factores de emisión que no corresponden a las características regionales, además del error muestral que caracteriza la naturaleza misma del fenómeno de estudio.

Teniendo en cuenta lo anterior, el dato de la emisión de metano por fermentación entérica tiene asociado un nivel de incertidumbre contenido en los factores de emisión utilizados y en los datos de la actividad, como se muestra a continuación:

- Incertidumbre factor de emisión.
- La incertidumbre del factor de emisión de metano por fer-

mentación entérica del IPCC 1996 es del 50% de acuerdo con lo reportado en las directrices del IPCC.

- Incertidumbre dato de actividad (*stock* de ganado bovino)
- La incertidumbre del dato de actividad está relacionada con los datos usados sobre la cantidad de ganado presente en el año a estimar la emisión, la cual se evalúa teniendo como guía conceptual los procedimientos propuestos por el IPCC y los elementos de análisis de incertidumbre del DANE.

4. Resultados

El total de las emisiones de metano por el *stock* de ganado bovino

es la suma de las emisiones en las subcategorías de ganado bovino (lechero y otro ganado).

En la tabla 4 se muestran los resultados de las emisiones para el periodo 2000- 2010 preliminar, utilizando el FE por defecto del IPCC 1996 que es el usado oficialmente por los Países en Desarrollo para presentar los inventarios nacionales de GEI en el contexto de la Convención Marco de las Naciones Unidas sobre Cambio Climático (CMNUCC); y en el gráfico 1, se representa el comportamiento creciente de las emisiones para este periodo.

A partir de los resultados de las emisiones del ganado bovino, se observa cómo las emisiones de metano crecen a medida que el *stock* de ganado bovino varía, existiendo una relación directa entre el número de animales y las emisiones.

Tabla 4. Emisiones de metano por la fermentación entérica derivada del ganado bovino del sector pecuario del SCN. 2000-2010p.

Año	Emisión de metano	Variación porcentual
	(Gg CH ₄ /año) FE-IPCC* 1996	
2000	1.157,65	
2001	1.179,51	1,89
2002	1.214,17	2,94
2003	1.248,38	2,82
2004	1.266,37	1,44
2005	1.290,33	1,89
2006	1.309,43	1,48
2007	1.334,56	1,92
2008	1.365,76	2,34
2009	1.397,72	2,34
2010p	1.423,71	1,86

Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

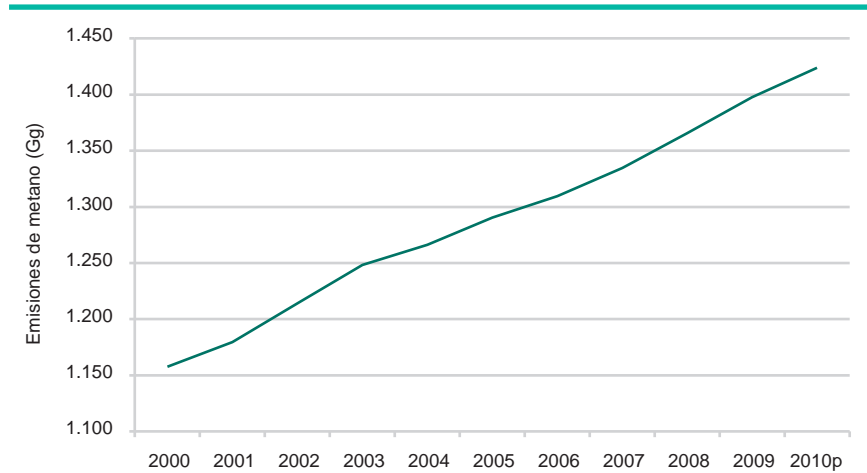
*FE-IPCC: factor de emisión (FE) propuesto por el IPCC en las Directrices del año 1996.

4.1. Indicador de intensidad de emisiones

En términos del indicador de intensidad de emisiones se debe tener en cuenta que actualmente no se encuentra detallado el PIB del ganado

bovino para cada año; por ello, se recurre a utilizar el PIB a precios constantes del total de la actividad 03 (Animales vivos, productos animales y productos de la caza) para cuantificar el indicador de intensidad de emisiones, que se obtiene

Gráfico 1. Emisiones de metano para el sector pecuario, ganado bovino. 2000-2010p.



Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

Tabla 5. Indicador de intensidad de emisiones. 2000-2010p.

Año	Emisiones de metano del stock de ganado bovino (toneladas)	Valor agregado de la actividad 03 a precios constantes del 2005 por encadenamiento	Indicador de intensidad de emisiones
		(miles de millones)	(toneladas/miles de millones de pesos)
2000	1.157.646,43	8.433,00	137,28
2001	1.179.506,93	8.545,00	138,03
2002	1.214.173,75	8.873,00	136,84
2003	1.248.377,00	9.068,00	137,67
2004	1.266.369,97	9.556,00	132,52
2005	1.290.328,78	9.919,00	130,09
2006	1.309.428,44	10.334,00	126,71
2007	1.334.556,77	10.700,00	124,72
2008	1.365.762,20	11.102,00	123,02
2009	1.397.716,34	11.008,00	126,97
2010p	1.423.710,00	11.122,00	128,01

Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

dividiendo las emisiones resultantes de la fermentación entérica del ganado bovino por el valor agregado a precios constantes de esta actividad 03 (tabla 5 y gráfico 2).

El indicador está relacionado con las emisiones de metano obtenidas utilizando FE IPCC 1996.

En el gráfico 2, se puede ver que el valor agregado de la actividad 03 a partir del 2003 ha tenido un creci-

miento mucho mayor al crecimiento en las emisiones de metano por fermentación entérica, y cuyo comportamiento se ha venido acercando desde el 2008 al de las emisiones.

El indicador nos muestra la relación en términos de emisiones de un gas en toneladas por cada mil millones de pesos del valor agregado de la actividad a precios constantes del 2005 por encadenamiento; por ejemplo, para el año

2010p por cada mil millones de pesos a precios constantes se emiten 128,01 toneladas de metano (gráfico 3). En efecto, el valor agregado a precios constantes de la actividad 03 (Animales vivos, productos animales y productos de la caza) crece durante este periodo en menor proporción que el crecimiento del *stock* de ganado, lo que implica una tendencia decreciente en la relación: emisiones/valor agregado.

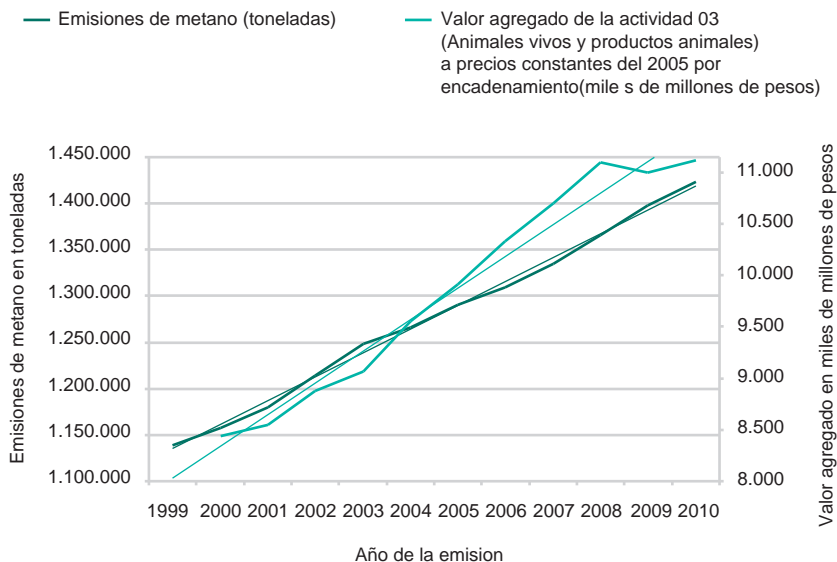
4.2. Emisiones en el ámbito nacional

Colombia, como parte de la CM-NUCC, hasta el momento ha presentado oficialmente los datos de emisiones de GEI para los años 1990, 1994, 2000 y 2004. A partir de esta información se estructura la comparación entre los datos del IDEAM en cuanto a la emisión de metano derivada por la fermentación entérica del ganado bovino y los resultados del piloto de la cuenta de emisiones para el periodo 2000-2010p (tabla 6).

De la tabla 6, no se puede derivar ningún análisis comparativo entre las fuentes de información. Las emisiones calculadas por el IDEAM tienen datos de actividad (*stock* del ganado bovino) de variadas fuentes de información como son el Instituto Colombiano Agropecuario (ICA), la Encuesta Nacional Agropecuaria (ENA) y la Federación Colombiana de Ganaderos (FEDEGAN). Los datos sobre el *stock* de ganado bovino utilizado en el piloto de la cuenta de emisiones del DANE provienen de las cifras referidas al inventario de ganado vacuno en cabezas, por grupo etario, según propósito y sexo calculado por la Dirección de Síntesis y Cuentas Nacionales como parte de la actividad pecuaria.

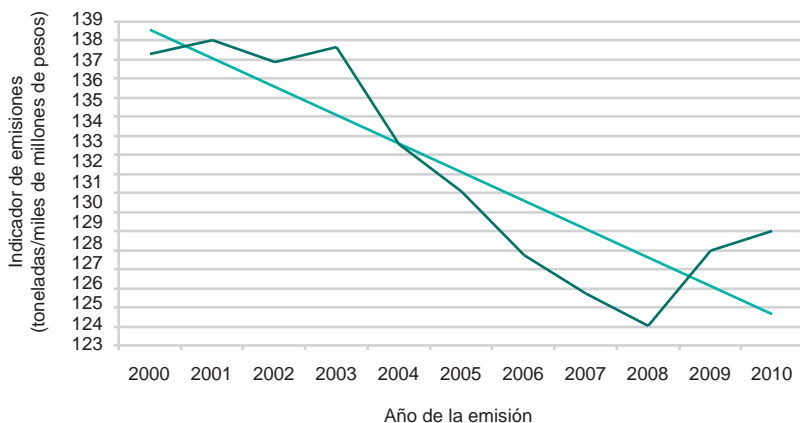
Por lo anterior, se evidencia la necesidad de fortalecer la medición de las emisiones de GEI del país, como insumo para los reportes periódicos ante la CMNUCC y para el segui-

Gráfico 2. Relación entre las emisiones de metano por fermentación entérica del ganado bovino y el valor agregado de la actividad 03 para el periodo 2000-2010p, utilizando el factor de emisión IPCC 1996.



Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

Gráfico 3. Indicador de intensidad de emisiones de metano por fermentación entérica del ganado bovino. 2000-2010p.



Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

*La línea verde muestra el indicador y la línea azul la tendencia histórica del indicador.

miento de la Estrategia Colombiana de Desarrollo Bajo en Carbono (ECDBC). Es importante coordinar los esfuerzos del IDEAM y del DANE para establecer la medición anual de

las emisiones de GEI en el ámbito de la cuenta de emisiones y de los Inventarios Nacionales de GEI.

4.3. Emisiones en el ámbito internacional

La CMNUCC, en su artículo 4, establece que los países deben reportar periódicamente las emisiones de GEI. Para el caso de los países que tienen compromisos de reducción de emisiones el reporte se hace anualmente; para los demás países (como Colombia) los informes de sus emisiones se estructuran y reportan dentro de las Comunicaciones Nacionales. Dentro de las emisiones que se deben presentar, están las emisiones de metano originadas por la fermentación entérica del ganado bovino. En las tablas 7 y 8 se muestran las emisiones en toneladas para los países Anexo I⁶, No Anexo I⁷ y Colombia (cálculo DANE, emisiones de metano con FE IPCC 1996), y mediante los cuales se establece la comparación en porcentaje.

El total de las emisiones de metano para cada año y para los grupos de países Anexo I y No Anexo I se suscribe a los datos que envían los países a la CMNUCC. La comparación se establece para el periodo 2000 a 2009 como se puede ver en la tabla 7; para el año 2010 todavía no existen reportes en la base de datos de la CMNUCC. En las columnas Colombia (%) vs. Anexo I y Colombia (%) vs. No Anexo I, se establece la proporción de la generación de emisiones del país frente al total de las emisiones de los Anexo I y No Anexo I.

En el total global, Colombia llega en el año 2009 a participar con el 3,43% de las emisiones globales de metano por fermentación entérica del ganado bovino (tabla 8). El crecimiento en la participación de las emisiones del país está directamente relacionado con el aumento

⁶ Los países Anexo I son los países industrializados que eran miembros de la OCDE (Organización para la Cooperación y el Desarrollo) en 1992, además de los países con economías en transición (PET), incluyendo la Federación de Rusia, los Estados Bálticos y de Europa Oriental y Central.

⁷ Los países No-Anexo I: son en su mayoría países en desarrollo. Algunos grupos de países en vías de desarrollo son reconocidos por la Convención como especialmente vulnerables a los efectos adversos del cambio climático, incluidos los países con zonas bajas costeras y aquellos propensos a la desertificación y la sequía. Otros (como los países que dependen en gran medida de los ingresos procedentes de la producción de combustibles fósiles y el comercio) son más vulnerables a los impactos económicos potenciales resultado de las medidas de respuesta al cambio climático.

en el *stock* de ganado, lo que genera una mayor participación en el total de emisiones mundiales.

Para el nivel de incertidumbre de los datos de actividad del periodo 1999-2009, por motivos de disponibilidad de información relacionada con los reportes del coeficiente de variación estimado (cve) a nivel de tipo de ganado, se presentan dos análisis de incertidumbre. El primero, que se puede observar en la tabla 9 para el periodo 2000-2006,

comprende el cálculo de incertidumbre desagregado por categoría de ganado bovino (leche y otro ganado); este procedimiento es el más acertado para obtener la incertidumbre de los datos de actividad, ya que existen marcadas diferencias en los cve por tipo de ganado.

En el segundo análisis, para el periodo 2007-2009 (tabla 10), se emplea el cve del total nacional, limitando el análisis de incertidumbre de los datos por tipo de ganado.

5. Información para la toma de decisiones

El quehacer del Departamento Administrativo Nacional de Estadística (DANE) está sustentado por el Decreto 262 de 2004, el cual establece como objetivo institucional «garantizar la producción, disponibilidad y calidad de la información estadística estratégica, y dirigir, planear, ejecutar, coordinar, regular y evaluar la producción y difusión de información oficial básica». Así mismo, la institución debe «elaborar las cuentas anuales, trimestrales, nacionales, regionales y satélites, para evaluar el crecimiento económico nacional, departamental y sectorial». En este sentido, en la Dirección de Síntesis y Cuentas Nacionales se desarrolla la Cuenta Satélite Ambiental (CSA) que tiene como objetivo general «medir en unidades físicas y monetarias, de forma sistémica y para cada período contable, la variación de los *stocks* de los activos ambientales, las interacciones entre el ambiente y la economía, dentro de la economía y de la economía al ambiente. De forma paralela y en coherencia con el Sistema de Cuentas Nacionales, la cuenta satélite mide el esfuerzo de los diferentes sectores económicos para conservar, mitigar o proteger el medio ambiente» (DANE, 2012: 17).

Igualmente, Colombia, a través del DANE, hace parte de la Comisión de Estadística de Naciones Unidas, que a partir del año 2008 inicia la labor de establecer cómo las estadísticas oficiales pueden proporcionar la información necesaria para el monitoreo, medición y análisis de los temas relacionados con el cambio climático, y asigna al Comité de Expertos sobre Cuentas Ambientales y Económicas de las Naciones Unidas (UNCEEA por sus siglas en inglés) la tarea de velar por la incorporación de la información sobre cambio climático en las estadísticas

Tabla 6. Comparación de las emisiones de metano (en toneladas) por fermentación entérica del ganado bovino, entre IDEAM y DANE. 2000-2010p.

Año de la emisión	Emisiones IDEAM	Emisiones DANE	Diferencia (toneladas)	Diferencia (%) IDEAM-DANE
	(toneladas CH ₄)	(toneladas CH ₄)	IDEAM-DANE	
2000	1.227.688,15	1.157.646,43	70.041,72	5,71
2001		1.179.506,93		
2002		1.214.173,75		
2003		1.248.377,00		
2004	1.263.321,18	1.266.369,97	-3.048,79	-0,24
2005		1.290.328,78		
2006		1.309.428,44		
2007		1.334.556,77		
2008		1.365.762,20		

Fuente: DANE-Dirección de Síntesis y Cuentas Nacionales.

Tabla 7. Comparación de las emisiones de metano (en toneladas) por fermentación entérica del ganado bovino, entre los países Anexo I, No Anexo I y Colombia (cálculos DANE – FE 1996). 2000-2009.

Año de emisión	Anexo I	No Anexo I	Colombia*	Colombia (%) vs. Anexo I	Colombia (%) vs. No Anexo I
2000	23.492.696,16	19.148.630,12	1.157.646,43	4,93	6,05
2001	23.451.521,55	19.083.963,11	1.179.506,93	5,03	6,18
2002	23.216.989,05	18.864.680,48	1.214.173,75	5,23	6,44
2003	22.964.829,52	18.754.827,14	1.248.377,00	5,44	6,66
2004	22.784.423,81	18.709.880,00	1.266.369,97	5,56	6,77
2005	22.838.490,95	18.864.792,86	1.290.328,78	5,65	6,84
2006	22.743.746,19	18.856.711,90	1.309.428,44	5,76	6,94
2007	22.707.007,62	18.835.379,52	1.334.556,77	5,88	7,09
2008	22.550.667,14	18.662.709,05	1.365.762,20	6,06	7,32
2009	22.300.804,13	18.448.193,36	1.397.716,34	6,27	7,58

Fuente: Base de datos de la CMNUCC.

*Datos calculados por la cuenta de emisiones DANE-Dirección de Síntesis y Cuentas Nacionales.

oficiales, con el fin de determinar los estándares y metodologías propias para su medición y análisis.

En este contexto, el DANE logró incorporar dentro del documento Conpes 3700 «Estrategia institucional para la articulación de políticas y acciones en materia de cambio climático en Colombia» los resultados de la cuenta de emisiones como parte de la información cuantitativa y cualitativa que da soporte a la toma de decisiones. Por ejemplo, en la implementación de la ECDBC, la cuenta de emisiones permite: i) identificar el aporte de

las actividades económicas en las emisiones de GEI del país, su evolución y relación con el crecimiento económico; ii) desarrollar opciones de mitigación propias para cada sector económico; y iii) puede ser útil como herramienta para el reporte y monitoreo de las emisiones anuales de GEI, por gas, por actividad económica y valor agregado.

Por otra parte, en el actual proceso de ingreso de Colombia a la Organización para la Cooperación y el Desarrollo Económico (OCDE), se requiere la producción de datos estadísticos de alta calidad

como medio para que los gobiernos comparen las experiencias de sus políticas, busquen respuesta a problemas comunes, identifiquen buenas prácticas y coordinen políticas domésticas e internacionales. En el contexto ambiental la OCDE ha puntualizado sus esfuerzos en el desarrollo de indicadores que permitan medir el desarrollo sostenible, entre los cuales –para el componente climático– se encuentra la construcción del indicador de emisiones de dióxido de carbono derivadas de las actividades económicas de un país.

Tabla 8. Comparación de las emisiones de metano (en toneladas) por fermentación entérica del ganado bovino, en el total global. 2000-2009.

Año de emisión	Anexo I + No Anexo I	Colombia*	Colombia-global (%)
2000	42.641.326,29	1.157.646,43	2,71
2001	42.535.484,66	1.179.506,93	2,77
2002	42.081.669,52	1.214.173,75	2,89
2003	41.719.656,67	1.248.377,00	2,99
2004	41.494.303,81	1.266.369,97	3,05
2005	41.703.283,81	1.290.328,78	3,09
2006	41.600.458,10	1.309.428,44	3,15
2007	41.542.387,14	1.334.556,77	3,21
2008	41.213.376,19	1.365.762,20	3,31
2009	40.748.997,49	1.397.716,34	3,43

Fuente: Base de datos de la CMNUCC.

*Datos calculados por la cuenta de emisiones DANE-Dirección de Síntesis y Cuentas Nacionales.

Tabla 10. Incertidumbre de los datos de actividad y del factor de emisión utilizados para la estimación de las emisiones de metano por fermentación entérica. 2007-2009.

Año	Coefficientes de variación para el total
2007	7
2008	4,1
2009	11,7

Fuente: Encuesta Nacional Agropecuaria.

Tabla 9. Incertidumbre de los datos de actividad, por categoría de ganado bovino utilizada para la estimación de las emisiones de metano por fermentación entérica. 2000-2006.

Año	Coeficiente de variación estimado por categoría de ganado bovino para la estimación de las emisiones de metano por fermentación entérica						
	Leche		Otro ganado				
	Leche hembras	Leche machos	Carne hembras	Carne machos	Doble propósito hembras	Doble propósito machos	Machos reproductores
2000	16,30	15,20	2,50	4,30	4,40	4,30	3,50
2001	8,20	13,40	2,20	3,60	2,40	4,30	3,30
2002	3,62	3,28	3,00	6,44	1,65	1,76	3,27
2003	15,40	13,00	2,90	6,40	2,30	3,90	3,80
2004	8,20	10,40	2,40	6,20	2,10	3,40	0,00
2005	8,60	11,30	2,50	6,40	2,20	3,60	0,00
2006	5,05	5,46	3,19	3,03	2,42	3,22	3,22

Fuente: Encuesta Nacional Agropecuaria.

Conclusiones

- La cuenta de emisiones permite cuantificar las emisiones anuales derivadas por las actividades económicas que integran las cuentas nacionales del país; es una herramienta útil para fortalecer el reporte periódico de emisiones en el contexto de los Inventarios Nacionales de GEI, de los requerimientos en el ámbito de la implementación de la ECDBC y de la generación de estadísticas oficiales asociadas al cambio climático.
- La relación en términos de emisiones de un gas por valor agregado de una actividad económica permite identificar y cuantificar el impacto del crecimiento económico en el medio ambiente, así como orientar en el ámbito de la política pública los mecanismos sectoriales para disminuir las emisiones.
- En el cálculo de las emisiones, al utilizar el factor de emisión por defecto del IPCC, se condiciona el dato de emisión a una alta incertidumbre. Es importante el desarrollo de factores propios que integren las circunstancias nacionales del país.
- La elaboración de la cuenta de emisiones debe responder a un proceso de arreglos institucionales que admitan la importancia de estructurar un sistema de registro de las emisiones al aire como parte de la generación de información adecuada, periódica y confiable para la toma de decisiones.
- La cuantificación de las emisiones de gases de efecto invernadero por actividad económica permite a los países establecer su línea base de emisiones e identificar las actividades sobre las cuales se deben orientar las estrategias de reducción de emisiones. Así, esta cuenta en los países de la comunidad europea ha permitido el seguimiento de las políticas de reducción de emisiones en el contexto de los compromisos adquiridos en el ámbito del protocolo de Kioto y en la generación anual de inventarios nacionales de GEI de alta calidad como parte de su responsabilidad ante la CMNUCC.

Bibliografía

- Bernal, Laila (s. f.). Efecto de la mezcla de las leguminosas *Calliandra closthyrsus*, *Flemingia macrophylla*, *Cratylia argentea* y *Vigna unguiculata* ensilada y henificada sobre los parámetros de fermentación ruminal in vitro y producción de leche en bovinos. Tesis de Maestría Ciencias Agrarias Producción Animal Tropical.
- Comisión Europea (Eurostat), Organización para la Agricultura y la Alimentación (FAO), Fondo Monetario Internacional (FMI), Organización para la Cooperación y el Desarrollo Económico (OCDE), Banco Mundial (BM) (2012). *System of Environmental Economic Accounting. Central Framework (SEEA)*. New York: United Nations.
- Departamento Administrativo Nacional de Estadística, DANE (2012). Metodología de la Cuenta Satélite Ambiental (CSA).
- European System for the Collection of Economic Information on the Environment (SERIEE) 1994 version. 2002 edition.
- Eurostat (2009). *Pocketbooks-Agricultural statistics*. (2009 edition). Main results-2007-08.
- Eurostat (2010). *Using official statistics to calculate greenhouse gas emissions. A statistical guide*.
- Eurostat (2011). Statistics in focus -10/2011- Driving forces behind EU-27 greenhouse gas emissions over the decade 1999-2008.
- IDEAM (2009). *Inventario Nacional de Fuentes y Sumideros de Gases de Efecto Invernadero 2000-2004*. Bogotá, D. C.
- IEEAF (2002). The European Framework for Integrated Environmental and Economic Accounting for Forests.
- Integrated Environmental and Economic Accounting (2003). *Handbook of National Accounting United Nations, European Commission, International Monetary Fund, Organization for Economic Co-operation and Development and World Bank*. Nueva York, 2003.

IPCC (1996). *Directrices del IPCC para los inventarios de gases de efecto invernadero* (versión revisada en 1996).

IPCC (2006). *Directrices del IPCC para los inventarios nacionales de gases de efecto invernadero* (versión revisada 2006). Vol. 1, Orien-

tación general y generación de informes, cap. 3, Incertidumbres, y vol. 4, cap. 10.

IPCC (2007). *Climate Change 2007: Synthesis Report*. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing

Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.

UNSC (2008). *System of National Accounts 2008*.

UNSD (2007). *System of environmental-economic accounting for water, SEEAW*. Final Draft.

Glosario

Acidificación: El término acidificación o, en términos comunes, «lluvia ácida», se refiere a uno de los fenómenos relacionados con el cambio global de la atmósfera, el cual se ha generado a partir de la interacción entre la emisión y formación de gases contaminantes, y la dinámica atmosférica. La lluvia ácida se caracteriza principalmente por la formación de precipitaciones con un carácter de acidez (que incluyen la lluvia, la nieve y la deposición de partículas), las cuales causan efectos importantes sobre el medio. Los principales precursores de la lluvia ácida son los óxidos de azufre y los óxidos de nitrógeno, compuestos químicos generados a través de dos vías principales: las fuentes antrópicas, principalmente a partir de la quema de combustibles fósiles, o las fuentes naturales, compuestas por emisiones volcánicas y fuentes termales, entre otras.

Calidad del aire: Hay cientos de contaminantes en el aire que se presentan en forma de partículas y gases. El material particulado está compuesto por pequeñas partículas líquidas o sólidas de polvo, humo, niebla y ceniza volante. Los gases incluyen sustancias como el monóxido de carbono, dióxido de azufre y compuestos orgánicos

volátiles. También se puede clasificar a los contaminantes como primarios o secundarios. Un contaminante primario es aquel que se emite a la atmósfera directamente de la fuente y mantiene la misma forma química, como por ejemplo, la ceniza de la quema de residuos sólidos. Un contaminante secundario es aquel que experimenta un cambio químico cuando llega a la atmósfera. Un ejemplo es el ozono que surge de los vapores orgánicos y óxidos de nitrógeno que emite una estación de gasolina o el escape de los automóviles. Los vapores orgánicos reaccionan con los óxidos de nitrógeno en presencia de luz solar y producen el ozono, componente primario del *smog fotoquímico*. Los contaminantes de aire también se han clasificado como contaminantes criterio y contaminantes no criterio. Los contaminantes criterio se han identificado como comunes y perjudiciales para la salud y el bienestar de los seres humanos. Se les llamó contaminantes criterio porque fueron objetos de estudios de evaluación publicados en documentos de criterios de calidad del aire.

Cambio climático: Cambio de clima atribuido directa o indirectamente a la actividad humana que altera la composición de la

atmósfera mundial y que se suma a la variabilidad natural del clima observada durante periodos de tiempo comparables (Convención Marco sobre el Cambio Climático [CMCC] de las Naciones Unidas, artículo 1, año 1992).

Cuentas satélite: Sistema de contabilidad adicional o paralelo que amplía la capacidad analítica de las cuentas nacionales. Su finalidad es evitar que se sobrecargue o se desorganice el sistema central. El Sistema de Contabilidad Ambiental y Económica Integrada (SCAEI) es un sistema satélite del Sistema de Cuentas Nacionales (SCN).

Compuestos orgánicos persistentes: Conocidos internacionalmente por sus siglas en inglés, POP (Persistent Organic Pollutants), son un conjunto de compuestos orgánicos fabricados artificialmente por el hombre. Aunque existen en forma natural, los cuales son muy tóxicos, pues tienen un tiempo de persistencia en el ambiente muy largo debido a sus características fisicoquímicas. Al ser un compuesto artificial, las bacterias, los procesos fotoquímicos, químicos y demás organismos no pueden descomponerlos y degradarlos fácilmente. Muchos tienen efectos

acumulativos, ya que se almacenan en los tejidos grasos fijándose en la cadena alimenticia y pueden tener efectos hormonales. Hacen parte de los compuestos orgánicos persistentes, los compuestos químicos tóxicos, que se clasifican en pesticidas, insecticidas organoclorados, herbicidas y PCB.

Gas de efecto invernadero, gas invernadero (GEI): Componente gaseoso de la atmósfera, natural o antropógeno, que absorbe y emite radiación en determinadas longitudes de onda del espectro de radiación infrarroja térmica emitida por la superficie de la Tierra, por la propia atmósfera y por las nubes. Esta propiedad da lugar al efecto invernadero. El vapor de agua (H₂O), el dióxido de carbono (CO₂), el óxido nitroso (N₂O), el metano (CH₄) y el Ozono (O₃) son los gases invernadero primarios de la atmósfera terrestre. La atmósfera contiene, además, cierto número de gases invernadero enteramente antropógenos, como los halocarbonos u otras sustancias que contienen cloro y bromo, contemplados en el Protocolo de Montreal. Además del CO₂, del N₂O y del CH₄, el Protocolo de Kyoto contempla los gases de efecto invernadero hexafluoruro de azufre (SF₆), los hidrofluorocarbonos (HFC) y los perfluorocarbonos (PFC) (IPCC, 2007).

Medio ambiente nacional y medio ambiente del resto del mundo: i) Medio ambiente nacional es el que está asociado con el territorio nacional, incluida el área marítima y submarina de uso exclusivo económico, y el espacio aéreo del país; ii) Medio ambiente del resto

del mundo: incluye las esferas ambientales de los países fuera del territorio nacional; iii) Medio ambiente (sistema ambiental): esfera total del medio ambiente.

Metano (CH₄): Es uno de los seis gases de efecto invernadero que el Protocolo de Kyoto se propone reducir. Es el componente principal del gas natural, y está asociado a todos los hidrocarburos utilizados como combustibles, a la ganadería y a la agricultura. El metano de estrato carbónico es el que se encuentra en las vetas de carbón (IPCC, 2007).

Metales pesados: Desde el punto de vista químico, los metales pesados están constituidos por elementos de transición y postransición incluyendo algunos metaloides como el arsénico y el selenio. Estos elementos tienen una gravedad específica significativamente superior a la del sodio, calcio, y otros metales ligeros. Estos elementos se presentan en diferente estado de oxidación en agua, aire y suelo y presentan diversos grados de reactividad, carga iónica y solubilidad en agua.

Principio de residencia: El concepto de residencia en las cuentas nacionales no se basa en criterios jurídicos o de nacionalidad. La residencia de cada unidad institucional es el territorio económico con el cual tiene la relación más fuerte; en otras palabras, su centro predominante de interés económico, que se visualiza en que realiza actividades en un período de tiempo prolongado –un año o más– (SCN 2008).

Precusores de ozono: Sustancias que producen el agotamiento del ozono estratosférico que se viene observando desde 1970 y se debe principalmente al aumento de las concentraciones de compuestos reactivos de cloro y bromo causado por la degradación de las sustancias agotadoras de ozono SAO antropógenas, entre ellas los halones, los CFC, los HCFC, el metilcloroformo (CH₃CCl₃), el tetracloruro de carbono (CCl₄) y el bromuro de metilo (CH₃Br).

Sistema ambiental: El sistema ambiental incluye todos los recursos naturales y los flujos de estos recursos que no se han integrado al sistema económico. Se pueden distinguir tres funciones básicas del medio ambiente: como recurso, sumidero y prestador de servicios. La cuenta ambiental de emisión al aire, más precisamente de la atmósfera, tiene como función primaria la de sumidero (o destino) de los residuos derivados de los flujos de las actividades de producción y consumo.

Sistema Económico: El sistema económico está constituido por un amplio número de entidades económicas, que realizan múltiples transacciones para lograr diferentes tipos de objetivos ya sea de producción, consumo, o acumulación. La unidad básica del sistema económico se denomina unidad institucional. Las unidades institucionales que conforman el Sistema Económico del país son aquellas que residen en el territorio económico, referidas como «unidades de residencia» en las cuentas nacionales.

Una comparación empírica entre algunos métodos demográficos y estadísticos para la estimación del tamaño de un área pequeña

Hugo Andrés Gutiérrez Rojas
José Fernando Zea Castro

Una comparación empírica entre algunos métodos demográficos y estadísticos para la estimación del tamaño de un área pequeña

Hugo Andrés Gutiérrez Rojas: Estadístico, magíster y estudiante del doctorado de Estadística de la Universidad Nacional de Colombia. Decano de la Facultad de Estadística de la Universidad Santo Tomás. Correo electrónico: hugogutierrez@usantotomas.edu.co

José Fernando Zea Castro: Estadístico de la Universidad Nacional, magíster en estadística de la misma universidad. Docente Investigador de la Facultad de Estadística de la Universidad Santo Tomás. Correo electrónico: josezea@usantotomas.edu.co

Fecha de recepción: 25 de julio de 2013

Fecha de aceptación: 15 de octubre de 2013

Resumen: uno de los mayores avances metodológicos en la teoría de muestreo se dio en la mitad del siglo pasado con la incorporación de los modelos probabilísticos de superpoblación en la inferencia de poblaciones finitas. Particularmente, estos métodos de muestreo son usados extensamente en la estimación de totales en subgrupos poblacionales con baja incidencia, comúnmente conocidos como dominios o áreas pequeñas (Brewer, 1999).

Paralelamente se han desarrollado varios métodos demográficos que apuntan a la estimación del total de una población en un dominio. Algunos de estos métodos no tienen en cuenta la naturaleza aleatoria de la estructura dinámica de la población, sino que se basan en teorías demográficas ampliamente aceptadas; entre estos se hallan las técnicas de tasas de vida (Bogue, 1950), los métodos compuestos (Bogue y Duncan, 1959) y el método de la unidad de vivienda (Smith y Lewis, 1980). Sin embargo, existen métodos demográficos que sí involucran una componente aleatoria de estimación; por ejemplo, los procedimientos de regresión sintomáticos –incluyendo el método de la razón de correlaciones (Schmitt y Crosetti, 1954) y el método de la diferencia de correlaciones (O’Hare, 1976)– y el método de regresión muestral (Ericksen, 1974).

La comparación empírica realizada mediante la simulación de una población grande en varios periodos de tiempo y subdividida en dominios mutuamente excluyentes arroja que, en términos de eficiencia en la estimación del total del dominio, el método estadístico más eficiente (que induce un menor error cuadrático medio) es el estimador sintético.

Palabras clave: áreas pequeñas, métodos demográficos, total poblacional, modelos de superpoblación.

Abstract: *a major methodological advance in survey sampling theory, proposed in the middle of last century, is the incorporation of super-population models in finite population inference. Particularly, these sampling methods are used extensively in estimating a population total from subgroups with low incidence, commonly known as domains or small áreas (Brewer, 1999).*

Alongside several demographic methods have been developed in order to estimate a population total from a domain. Some of these methods do not take into account the random nature of the dynamic structure of the population, but are based on widely accepted demographic theories, among these techniques are rates of living (Bogue, 1950), the composite methods (Bogue & Duncan, 1959) and the housing unit type (Smith & Lewis, 1980). However, there are demographic methods that do involve a random component in the estimation, e.g. symptomatic regression procedures –including the correlation ratio method (Schmitt & Crosetti, 1954) and the correlation difference method (O’Hare, 1976)– and the sample regression method (Ericksen, 1974).

The empirical comparison is done by means of simulating a large population in various time periods and subdivided it into mutually exclusive domains. This exercise shows that, in terms of efficiency in the estimation of a domain’s total, the sintetic estimator provieds lower MSE.

Key Words: *demographic methods, population total, small areas, superpopulation models.*

1. Introducción

El término área pequeña es utilizado para denotar un área geográfica determinada. Por ejemplo, un área pequeña puede ser un condado, una municipalidad, un departamento o cualquier división censal. Sin embargo, esta definición no está supeeditada únicamente a la acepción geográfica. Un dominio o un área pequeña puede estar relacionado con una subpoblación pequeña conformada por elementos dentro de un grupo específico como edad-género-nivel socioeconómico.

En este trabajo de investigación seguiremos el enfoque de la subcomisión en muestreo de las Naciones Unidas, la cual define a un dominio como un subgrupo poblacional que cumple las siguientes características (Gutiérrez, 2009):

- a. Todo elemento de la población debe pertenecer a un dominio.
- b. Ningún elemento de la población puede pertenecer a dos dominios.
- c. La reunión de todos los dominios es la población de interés.

Además de lo anterior, un dominio es un subgrupo poblacional de interés; es decir, que se requieren estimaciones precisas para cada subgrupo de interés en un estudio o en la selección de una muestra. Sin embargo, una limitación importante, en términos de estimación eficiente, son los recursos económicos y las dificultades logísticas propias de la selección de una muestra. Por tanto, si el tamaño de la muestra en el estudio general no es amplio, se corre el riesgo de que la cobertura muestral para cada uno de estos dominios de interés sea muy baja o incluso nula. El anterior razonamiento da pie a la definición formal de dominio pequeño. Rao (2003, p. 1) define un área (dominio) grande si el tamaño de muestra en ese do-

minio específico es suficientemente grande para generar estimaciones directas¹ con un buen nivel de precisión; por otro lado, se define un área (dominio) pequeña como un subgrupo para el cual no hay un número suficiente de individuos en la muestra y no se permiten estimaciones directas puesto que el nivel de precisión es bajo.

La vida práctica del investigador está llena de situaciones en donde se necesitan estimaciones de los totales poblacionales de dominios pequeños. Por ejemplo, una encuesta nacional de educación arrojará estimaciones muy precisas acerca del total de niños con problemas de aprendizaje en el país. Sin embargo, cuando se trata de hallar ese total discriminado por departamentos, ciudades o barrio, la precisión de las estimaciones en la encuesta decrece sustancialmente puesto que el tamaño de muestra será pequeño. Por esta razón es necesario acudir a otros recursos para encontrar estimaciones precisas que den cuenta de los parámetros de interés en las áreas pequeñas. De esta forma, existen métodos demográficos

¹ La estimación directa en un dominio se refiere a la inferencia estadística surgida de métodos tradicionales como, por ejemplo, la inferencia basada en el diseño de muestreo del estudio, o también a la utilización de ponderaciones para cada individuo de tal forma que representen a la población. La estimación indirecta se refiere al uso de modelos estadísticos predictivos que describen la muestra. En este enfoque el objetivo es predecir el valor de la característica de interés para los individuos que no fueron incluidos en la muestra y se supone que el diseño de muestreo es no informativo para la predicción.

y métodos estadísticos que pueden coadyuvar al investigador a obtener dichas estimaciones.

Con respecto a los métodos de muestreo estadístico tradicional, vale la pena aclarar que estos procedimientos son conocidos en la literatura como métodos de estimación directa. En relación con lo anterior, Gutiérrez (2009) afirma que los métodos directos de estimación realizan las estimaciones de los parámetros dependiendo del diseño de muestreo escogido para seleccionar la muestra sin tener en cuenta las propiedades de la población finita. Así, por ejemplo, un estimador del total poblacional de alguna característica de interés y , t_y , es

$$\hat{t}_y = \sum_{k \in S} d_k y_k$$

En donde d_k es una ponderación inducida por el diseño de muestreo. Bajo esta perspectiva, los valores y_k son tomados como la observación en el individuo k de la característica de interés y . Sin embargo, no se toma como una variable aleatoria, sino como una cantidad fija.

Por otro lado, los métodos indirectos suponen uso de información auxiliar relacionada a la característica de interés mediante un modelo de superpoblación ξ . Bajo esta perspectiva no se requiere que los datos provengan de una muestra probabilística (la forma en la que se escoge la muestra no se tiene en cuenta para la estimación de los parámetros de interés), si el diseño de muestreo se supone no informativo, y la observación de la característica de interés en las unidades poblacionales y_k se define como la realización de una variable aleatoria Y_k . De esta forma, el total poblacional se puede escribir como

$$T_y = \sum_{k \in S} Y_k + \sum_{k \notin S} Y_k$$

Y su predicción estará dada por

$$\hat{T}_y = \sum_{k \in S} Y_k + \sum_{k \notin S} E_{\xi}(Y_k)$$

En donde $E_{\xi}(Y_k)$ es una predicción del valor Y_k realizada con los datos obtenidos de la muestra seleccionada s . De esta manera, Brewer (2002) afirma que no es común utilizar estimaciones indirectas en dominios grandes puesto que el modelo supuesto puede ser errado y tampoco es común utilizar estimaciones directas en dominios pequeños puesto que el tamaño de muestra es pequeño y no es adecuado utilizar el principio de representatividad.

La organización del artículo es como sigue: la sección 1 describe con detalle los subgrupos poblacionales, la diferenciación con áreas pequeñas y los métodos de estimación. La sección 2 muestra algunas técnicas demográficas y estadísticas que son ampliamente utilizadas para la estimación de totales poblacionales en áreas pequeñas. La sección 3 está dedicada al estudio empírico de tales métodos de estimación, presenta la estructura de la simulación y concluye que el mejor método de estimación es aquel que modela la estructura aleatoria en la muestra seleccionada. En la sección 4 se recrea un escenario real que describe paso a paso el uso de estos métodos para la estimación del tamaño de un departamento. Finalmente, en la última sección se dan algunas conclusiones junto con una pequeña discusión.

2. Métodos de estimación

El trabajo de estimar el total poblacional en un área pequeña surgió hace varias décadas aunque con el pasar del tiempo los métodos que se utilizan son cada vez más refinados. En esta sección se revisan algunos de estos métodos tanto del

punto de vista demográfico como desde el punto de vista estadístico mediante la incorporación de estimaciones indirectas surgidas de modelos de superpoblación que no están ligados al diseño de muestreo.

2.1. Métodos demográficos

Rao (p. 27, 2003) afirma que los métodos demográficos utilizados para la estimación en áreas pequeñas emplean estimadores indirectos basados en el vínculo de modelos explícitos. Sin embargo, estos no son modelos que describen el comportamiento de la población a nivel de cada individuo, sino que son modelos tradicionales, en algunos casos surgidos directamente de las teorías de la economía, que no tienen en cuenta un término aleatorio y casi siempre son determinísticos. Adicionalmente con cada uno de estos métodos no es posible realizar una medición de la incertidumbre de las estimaciones puesto que no hay un componente aleatorio. A continuación se hace un recuento de algunas de ellas para la estimación del total poblacional de un área pequeña en un tiempo t , denotado como P_t .

2.1.1. Tasas de vida

Propuesto por Bogue en 1950, utiliza los nacimientos y las defunciones en el tiempo t y asume que se conoce una estimación del total poblacional de un dominio grande, que contiene al dominio (pequeño) de interés, en el tiempo y denotada como P_t . De esta manera, la estimación del total poblacional para el dominio de interés está dada por

$$\hat{p}_t = \frac{1}{2} \left(\frac{b_t}{\hat{r}_{1t}} + \frac{d_t}{\hat{r}_{2t}} \right) \quad (1)$$

Donde b_t y d_t denotan el número de nacimientos y muertes en el dominio de interés, que se suponen conocidos mediante registros oficiales. Además, para cada caso se supone que la tasa cruda de natalidad, r_{1t} , es directamente proporcional a

la tasa cruda de natalidad conocida para el último censo, r_{10} . Análogamente, se supone que la tasa cruda de mortalidad, r_{2t} , es directamente proporcional a la tasa cruda de natalidad conocida para el último censo, r_{20} . Por tanto,

$$\hat{r}_{it} = \hat{\phi}_i r_{i0} \quad (i = 1,2)$$

Donde los factores multiplicativos $\hat{\phi}_1$ y $\hat{\phi}_2$ están dados por

$$\hat{\phi}_1 = \frac{\hat{R}_{1t}}{R_{10}} \text{ y } \hat{\phi}_2 = \frac{\hat{R}_{2t}}{R_{20}}$$

Respectivamente, donde R_{10} denota la tasa cruda de natalidad del último censo en el dominio grande y R_{20} denota la tasa cruda de mortalidad del último censo en el dominio grande. Por último, $\hat{R}_{1t} = B_t/\hat{P}_t$ y $\hat{R}_{2t} = D_t/\hat{P}_t$ y denotan las tasas crudas estimadas de natalidad y mortalidad en el tiempo t . Un supuesto fuerte de este método es que la relación de proporcionalidad se considera igual tanto en el dominio grande como en el dominio de interés.

2.1.2. Métodos compuestos

Rao (2003) afirma que esta familia de métodos calcula, mediante la técnica anterior, estimaciones separadas de subgrupos poblacionales específicos para luego sumarlas y obtener una estimación compuesta de P_t . En particular, el método de componentes propuesto por Bogue y Duncan (1959) supone el conocimiento de información auxiliar censal para un periodo anterior. De esta forma, la estimación del total poblacional del dominio de interés está dada por

$$\hat{p}_t = p_0 + b_{0,t} - d_{0,t} + m_{0,t} \quad (2)$$

Donde p_0 denota la población del dominio de interés en el tiempo

censal anterior, $b_{0,t}, d_{0,t}, m_{0,t}$ denotan el número de nacimientos, muertes y migración neta en el periodo $[0, t]$, respectivamente. Nótese que la migración neta puede ser positiva o negativa.

$$\hat{p}_t = (h_t)(pph_t) + gq_t \quad (3)$$

2.1.3. Métodos de la unidad de vivienda

Este tipo de métodos considera cualquier lugar de habitación, diferente al hogar, literalmente como un lugar de residencia; por ejemplo prisiones o residencias académicas. A estos lugares de residencia se les conoce con el nombre de lugares de alojamiento. De esta manera, una estimación para el parámetro de interés es la suma del total de personas en las unidades de vivienda (hogares) junto con el número de personas residiendo en lugares de alojamiento y está dado por

Donde $h(t)$ es el número de unidades de vivienda ocupadas en el tiempo t , pph_t , corresponde al número promedio de personas por hogar en el tiempo t y por último gq_t representa el número de personas en lugares de alojamiento en el tiempo t . Smith y Lewis (1980) proveen las expresiones necesarias para estimar las anteriores cantidades.

2.1.4. Métodos de razón y diferencia de correlaciones

Estudiado por Schmitt y Crosetti (1954), este método forma parte de los métodos de regresión que utilizan indicadores sintomáticos² para ajustar un modelo de regresión múltiple para estimar el total poblacional del dominio de interés. Así, es necesario disponer de información censal de dos periodos anteriores para ajustar un modelo de regresión del cambio en proporción, de un periodo censal a otro, de los valores de una variable dependiente contra las variables sintomáticas auxiliares.

Para el método de razón de correlaciones, se define U_i como el cambio en proporción del tamaño poblacional del i -ésimo dominio de interés en los dos periodos censales anteriores y Z_{ij} el cambio en proporción de la j -ésima variable sintomática en el i -ésimo dominio de interés en los dos periodos censales anteriores. De esta forma, el siguiente modelo de regresión puede ser ajustado mediante técnicas de estimación como la de mínimos cuadrados o la de máxima verosimilitud.

$$U_i = \gamma_0 + \gamma_1 z_{i1} + \dots + \gamma_p z_{ip} + u_i$$

En donde U_i son errores no correlacionados y las siguientes cantidades son conocidas para los periodos censales anteriores:

$$U_i = \frac{p_{i1}/P_1}{p_{i0}/P_0} \text{ y } z_{ij} = \frac{s_{ij1}/S_{j1}}{s_{ij0}/S_{j0}}$$

Para el i -ésimo dominio ($i=1, \dots, D$) y la j -ésima variable sintomática ($j=1, \dots, P$). Nótese que p_{i1} es el tamaño del i -ésimo dominio en el último periodo censal P_{i0} , es el tamaño del i -ésimo dominio en el primer periodo censal, $p_0 = \sum_i p_{i0}$, $P_1 = \sum_i p_{i1}$, S_{ij1} es el valor de la j -ésima variable sintomática para el i -ésimo dominio en el último periodo censal, S_{ij0} es el valor de la j -ésima variable sintomática para el i -ésimo dominio en el primer periodo censal, $s_{j0} = \sum_i s_{ij0}$ y $s_{j1} = \sum_i s_{ij1}$. Una vez que se encuentran las estimaciones de los parámetros de la regresión, $(\hat{\gamma}_0, \dots, \hat{\gamma}_p)$, se procede a predecir los valores de cambio para el tiempo actual t mediante la siguiente expresión.

$$Y^*_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \dots + \hat{\gamma}_p x_{ip}$$

En donde

$$Y^*_i = \frac{p_{it}/P_t}{p_{i1}/P_1} \text{ y } x_{ij} = \frac{s_{ijt}/S_{jt}}{s_{ij1}/S_{j1}}$$

Nótese que Y^*_i no se conoce y se tiene que predecir, aunque x_{ij} sí se conoce para todos los dominios de la población. De esta forma, el

² Un indicador sintomático se refiere a registros administrativos que contienen datos actualizados acerca de variables demográficas cuyos cambios están altamente correlacionados con los cambios en los dominios de interés. Por ejemplo, los cambios en el número de nacimientos, defunciones o en la migración (Rao, 2003).

total poblacional para el i -ésimo dominio de interés está dado por

$$\hat{p}_{it} = \left[Y^*_i - \left(\frac{p_{i1}}{P_1} \right) \right] \hat{P}_t \quad (5)$$

Y se supone que \hat{P}_t es una estimación acertada de $P_t = \sum p_{it}$ disponible de otras fuentes administrativas. Por otro lado, el método de diferencia de correlaciones (O'Hare, 1976) asume que se sabe la diferencia de las cantidades anteriores; es decir, se conoce:

$$U_i = p_{i1}/P_1 - p_{i0}/P_0 \text{ y } z_{ij} = s_{ij1}/S_{j1} - s_{ij0}/S_{j0}$$

Cuando se realiza la regresión entre las anteriores cantidades se obtienen estimaciones para los parámetros ($\hat{\gamma}_0, \dots, \hat{\gamma}_p$) y se procede a predecir los valores de cambio para el tiempo actual t mediante la siguiente expresión.

$$Y^*_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \dots + \hat{\gamma}_p x_{ip}$$

En donde

$$Y^*_i = p_{it}/P_t - p_{i1}/P_1 \text{ y } x_{ij} = s_{ijt}/S_{jt} - s_{i\Box 1}/S_{j1}$$

Una vez más, Y^*_i no se conoce y se debe predecir, aunque sí se conoce para todos los dominios de la población. De esta forma, el total poblacional para el i -ésimo dominio de interés está dado por

$$\hat{p}_{it} = \left[Y^*_i - \left(\frac{p_{i1}}{P_1} \right) \right] \hat{P}_t \quad (5)$$

2.1.5. Método de regresión muestral

En el anterior apartado, el cambio en la población se pretendía predecir mediante el uso de una expresión de regresión en donde los parámetros estaban estimados usando variables dependientes de los dos últimos censos. Sin embargo, si la diferencia poblacional entre el tiempo actual y los tiempos censales es significativa, entonces estas predicciones estarán sesga-

das. Ericksen (1974) propuso estimar los coeficientes de la regresión de una manera distinta al suponer que mediante una muestra se pueden conseguir estimaciones insesgadas del total poblacional de algunos dominios de la población, \hat{p}_{it} , y que las variables sintomáticas s_{ijt} se encuentran disponibles para todos los dominios de interés en el tiempo t . De esta manera, para los dominios que tienen estimaciones insesgadas se propone la siguiente regresión

$$\hat{p}_{it} = \beta_0 + \beta_1 s_{i1t} + \dots + \beta_p s_{ipt} + v_{it}$$

Como lo afirma Olayiwola (1986), la ecuación de regresión se aplica directamente para producir predicciones y/o estimaciones de los tamaños poblacionales de los dominios de interés. Por tanto, una estimación del tamaño poblacional para el i -ésimo dominio está dada por

$$\hat{p}_{it} = \hat{\beta}_0 + \hat{\beta}_1 s_{i1t} + \dots + \hat{\beta}_p s_{ipt} \quad (6)$$

2.2. Métodos estadísticos

En este apartado se pretenden recolectar algunas formas de estimación de los totales poblacionales para un dominio de interés. Como se afirmó en la introducción de este artículo, cuando el dominio es un área pequeña (o cuando se requieren estimaciones en dominios para los cuales la encuesta no fue diseñada) la mayoría de procedimientos obvian los pesos de muestreo, puesto que utilizarlos conllevaría a estimaciones con muy altos errores cuadráticos medios que se hacen inaceptables en términos de inferencia estadística, y utilizan modelos poblacionales que describan el comportamiento dentro del subgrupo poblacional de interés.

2.2.1. Estimación directa

Este tipo de estimadores utilizan los pesos resultantes del diseño de muestreo como principio de estimación. El espíritu que hay detrás de la estimación directa es que

cada uno de los individuos seleccionados en la muestra se representa a sí mismo y a un subconjunto de la población. Este tipo de procedimientos clásicos van muy bien cuando el tamaño de muestra es grande. Sin embargo, bajo este contexto, cuando se trata de un dominio pequeño, el principio de representatividad es muy susceptible puesto que con una muestra demasiado pequeña los errores de estimación serán demasiado altos. La familia de estimadores de Horvitz-Thompson, el estimador de Hájek (si no existe información auxiliar), el estimador general de regresión, el estimador de post-estratificación y el estimador de calibración (si existe información auxiliar) son casos particulares de estimadores directos cuya base inferencial es el diseño de muestreo que fue utilizado para seleccionar la muestra (así sea una muy pequeña) en cada dominio (Gutiérrez, 2009). Por ejemplo, si existe una variable de información auxiliar para todos los dominios, entonces un estimador directo, denominado de razón, para el tamaño poblacional del i -ésimo dominio sería

$$\hat{p}_i = t_{xi} \frac{\hat{t}_{yi}}{\hat{t}_{xi}} \quad (7)$$

Donde es el total conocido de la variable auxiliar para el i -ésimo dominio de interés y \hat{t}_{xi} , \hat{t}_{yi} corresponden a estimaciones de Horvitz-Thompson para el total de la característica de interés y del total poblacional en el i -ésimo dominio de interés, respectivamente.

2.2.2. Estimación sintética

Rao (2003) afirma que un estimador sintético es un estimador directo que tiene la característica de ser preciso para un dominio grande, que contiene al dominio de interés, pero que es aplicado al dominio pequeño al suponer que se comporta igual que el dominio grande. Si existe una variable de información auxiliar para todos los dominios,

entonces un estimador de razón dado para el total poblacional, t_y , estaría dado por

$$\hat{t}_y = t_x \frac{\hat{t}_y}{\hat{t}_x}$$

Ahora, si se asume que el dominio de interés se comporta de manera similar a la población completa, entonces un estimador sintético de razón para el tamaño poblacional del i -ésimo dominio sería

$$\hat{p}_i = t_{xi} \frac{\hat{t}_y}{\hat{t}_x} \quad (8)$$

2.2.3. Estimación compuesta

Si el supuesto sobre el que se construye el estimador sintético es demasiado débil, entonces habrá un sesgo considerable en la estimación resultante. Por esto, en algunas ocasiones, es mejor recurrir a un promedio ponderado entre un estimador sintético y un estimador directo. De esta forma, Rao (2003) afirma que se evade la inestabilidad del estimador sintético a la vez que se disminuye el sesgo del estimador directo. Por lo tanto, un estimador compuesto para el tamaño poblacional del i -ésimo dominio de interés está dado por

$$\hat{p}_{ic} = \varphi_i \hat{p}_{i1} + (1 - \varphi_i) \hat{p}_{i2} \quad (9)$$

En donde la ponderación φ_i ($0 < \varphi_i < 1$). Schaible (1978) afirma que una forma de estimar óp-

timamente a φ_i es mediante la siguiente expresión

$$\hat{\varphi}_i \approx \frac{ECM_d(\hat{p}_{i2})}{ECM_d(\hat{p}_{i1}) + ECM_d(\hat{p}_{i2})}$$

En donde $ECM_d(\hat{p}_{i1})$ representa el error cuadrático medio basado en el diseño de muestreo del estimador \hat{p}_{i1} .

3. Comparación empírica

En esta sección se presenta un conjunto de simulaciones que permiten tener un acercamiento al comportamiento estructural, en términos de eficiencia y sesgo, de algunos de los métodos demográficos y estadísticos. Para esto, se simulan dos escenarios, dependiendo del método de estimación utilizado. Los procedimientos demográficos que se tendrán en cuenta son: el método de la razón de correlaciones, el método de la diferencia de correlaciones y el método de la regresión muestral cuando se tiene acceso a una variable sintomática. Por otro lado, los procedimientos estadísticos involucrados en este estudio empírico son el estimador directo de expansión, y el estimador sintético de razón cuando se tiene acceso a una variable de totales auxiliares para los dominios.

se seleccionó una muestra aleatoria simple de tamaño $n=100$ y en cada dominio se seleccionó una muestra de 10 elementos.

El primer escenario del ejercicio empírico trata con los métodos demográficos. Para esta configuración, los tamaños de los dos periodos censales anteriores y del periodo actual se muestran en la tabla 1. Nótese que los parámetros de interés son el total poblacional igual a 5000, el total poblacional del primer dominio igual a 2000, hasta el total poblacional del primer dominio igual a 100. Como variable sintomática, que debe ser un indicador de los cambios en la población, para los tres periodos y en cada dominio se generaron números aleatorios provenientes de una distribución normal con media $\sqrt{p_{ia}}$ y varianza la unidad para cada $a=0,1,t$. Los valores resultantes de las variables sintomáticas y la suma por periodo se muestran en la tabla 2.

Para este ejercicio, el tamaño poblacional $P_t = 5000$ se supuso conocido y las estimaciones resultantes se muestran en la tabla 3. Por otro lado, para la simulación del procedimiento de regresión muestral, se supuso una única encuesta de tamaño 100 (10 entrevistas por dominio) y se utilizó el estimador directo de razón con una variable auxiliar generada para cada individuo en la población de interés con distribución normal de media 10y varianza $\sqrt{2}$. Una vez que el estimador de razón arrojó sus resultados, estos se utilizaron como variable dependiente en una regresión con intercepto contra como variable independiente y las estimaciones de este modelo lineal son las propias del procedimiento de regresión muestral cuyos resultados se muestran en la tabla 3.

Tabla 1. Descripción de la población en los dos periodos censales, en el periodo actual y en la muestra.

Dominio	1	2	3	4	5	6	7	8	9	10	Total
Censo 0=p_0	1810	810	310	310	45	45	45	45	30	30	3480
Censo 0=p_1	1875	875	375	375	75	75	75	75	60	60	3920
Tamaño=p_it	2000	1000	500	500	200	200	200	200	100	100	5000

Fuente: elaboración propia.

En cada uno de los escenarios se simuló una población finita de tamaño $N=5000$ la cual está dividida en 10 subgrupos poblacionales

para los que se requieren estimaciones específicas acerca de su tamaño poblacional que está dado en la tabla 1. En cada simulación

Nótese que bajo el contexto de procedimientos demográficos todos los estimadores del tamaño del dominio no son aleatorios sino que se

Tabla 2. Variables sintomáticas en los dos periodos censales y en el periodo actual.

Variable	1	2	3	4	5	6	7	8	9	10	S
s_i0	42,25	29,55	17,63	14,45	6,97	4,71	8,11	6,98	3,47	6,49	140,60
s_i1	44,79	28,21	18,20	19,73	9,29	11,33	9,99	7,92	5,05	8,76	163,28
s_1t	44,66	31,56	22,23	23,87	13,24	13,87	15,12	14,07	9,51	10,45	198,57

Fuente: elaboración propia.

refieren a cantidades fijas provenientes de expresiones determinísticas. Sin embargo, en términos de desviación relativa, el procedimiento que arrojó mejores resultados es el de regresión muestral, aunque se debe destacar que son peores que los que arrojó el estimador directo de razón que no corresponde a un procedimiento demográfico.

El segundo escenario se refiere a la comparación de estimadores estadísticos que sí involucran aleatoriedad para distintas muestras. Se comparan tres estimadores: el estimador directo

de razón con la misma variable auxiliar generada en el anterior escenario, el estimador sintético de razón y recurriendo a la simulación empírica se calcula un factor óptimo para el cálculo de un estimador compuesto que será función de los anteriores estimadores. Para cada muestra seleccionada se incluyeron diez elementos de cada dominio de interés. Este proceso se repitió $nsim=1000$ veces y los estimadores se compararon con respecto al sesgo relativo y al error cuadrático medio relativo. La tabla 4 muestra los resultados obtenidos para esta configuración.

Nótese que el sesgo de los estimadores directos y sintéticos es casi despreciable para la mayoría de los dominios de interés. Sin embargo, el estimador compuesto siempre tiene un sesgo mayor que los otros estimadores. Por otro lado, el error cuadrático medio del estimador sintético es mucho menor que el de los otros estimadores; aunque la variación del estimador compuesto es mucho mayor que la de los otros estimadores. Bajo estas condiciones el estimador sintético presenta una buena caracterización para la estimación del total poblacional en los dominios pequeños. Lo anterior se debe a que en la simulación empírica se respetó el hecho de que existiese una relación parecida entre la razón de los totales poblacionales de cada dominio con respecto a la razón de los totales de la población finita.

Tabla 3. Estimaciones y desviación relativa para el tamaño de cada dominio utilizando los métodos de razón y diferencia de correlaciones, un estimador directo y el método de regresión muestral.

Estimación	1	2	3	4	5	6	7	8	9	10	Total
p_r,i	2930,5	1409,9	619,5	617,9	130,0	124,0	132,6	140,5	115,0	98,4	6318,3
Desviación	47%	41%	24%	24%	-35%	-38%	-34%	-30%	15%	-2%	26%
p_d,i	2504,6	1147,7	477,3	479,7	73,4	94,6	61,5	44,6	37,7	78,9	5000,0
Desviación	25%	15%	-5%	-4%	-63%	-53%	-69%	-78%	-62%	-21%	0%
p_directa,i	1764,1	1048,4	478,3	533,6	195,4	230,2	206,7	196,3	100,7	98,4	4852,1
Desviación	-11%	4,8%	-4,3%	6,7%	-2,3%	15,1%	3,4%	-1,9%	0,7%	-1,6%	-3,0%
p_rm,i	1633,1	1113,5	647,8	601,0	140,8	199,3	212,4	230,7	52,0	21,5	4852,1
Desviación	-18%	11%	30%	20%	-30%	0%	6%	15%	-48%	-79%	-3,0%

Fuente: elaboración propia.

4. Aplicación real en el departamento del Chocó

Suponga que se requiere la estimación del total poblacional para un área o dominio pequeño. En esta sección se utilizarán los datos de la Gran Encuesta Integrada de Hogares para el periodo de diciembre del año de 2011. Debido a que la información anonimizada que provee

el DANE en su portal web presenta como nivel mínimo de georreferenciación al departamento, se decidió trabajar con el departamento como área pequeña. En algunos métodos se considera la información del último censo o incluso la de los dos últimos censos. En los años en los cuales no esté disponible el total de población se utiliza la información de las proyecciones de población. Algunas variables sintomáticas

como el número de nacimientos (registro de estadísticas vitales del DANE) y el total poblacional por sexo son consideradas en este ejercicio, gracias a que mantienen una alta correlación con la variable de interés en el dominio de interés (departamento del Chocó).

Tasas de vida: El área grande considerada es el territorio nacional (Colombia) y se asume con

un comportamiento similar al del área pequeña (departamento de Chocó). Para esta área pequeña, el número de nacimientos en el año 2011 fue de $b_t = 5587$, el número de muertes (no fetales) fue calculado en $d_t = 998$. Los factores multiplicativos se estimaron como

$$\hat{\phi}_1 = \frac{\hat{R}_{1t}}{R_{10}} = \frac{0,014453}{0,016786} = 0,86 \text{ y } \hat{\phi}_2 = \frac{\hat{R}_{2t}}{R_{20}} = \frac{0,004253}{0,004407} = 0,96$$

Por lo anterior, se estimó que $\hat{r}_{1t} = \hat{\phi}_1 r_{10} = (0,86)(0,011411) = 0,009824$ y $\hat{r}_{2t} = \hat{\phi}_2 r_{20} = (0,96)(0,002387) = 0,002304$. Se obtiene una estimación del total poblacional para el dominio de interés:

$$\hat{p}_{Chocó} = \frac{1}{2} \left(\frac{5587}{0,009824} + \frac{998}{0,002304} \right) = 500.948$$

Método compuesto: El número de nacimientos en Chocó entre el

periodo censal y 2011 se calculó en $b_{0,t} = 36144$, el número de defunciones entre 2005 y 2011 se calculó en $d_{0,t} = 7784$ y el saldo neto migratorio para Chocó para los años 2005-2011 se estimó³ en $m_{0,t} = -40588$ obteniéndose como estimación final para el total poblacional del dominio de interés: $\hat{p}_t = p_0 + b_{0,t} - d_{0,t} + m_{0,t} = 454.053 + 36.144 - 7.784 - 40.588 = 441.825$.

Tabla 4. Sesgo relativo y error cuadrático medio para la estimación del tamaño de cada dominio utilizando estimadores directos, sintéticos y compuestos.

Estimación	1	2	3	4	5	6	7	8	9	10
Sesgo (directo)	0,05	0,55	0,59	0,58	0,33	0,54	0,66	0,45	0,25	0,50
ECM (directo)	16746,23	4424,64	1046,76	1172,42	171,32	140,22	207,98	123,44	32,54	41,35
Sesgo (sintético)	-0,06	0,15	-1,48	1,07	-1,36	-0,11	1,15	1,40	3,13	-2,96
ECM (sintético)	1683,38	424,51	157,27	136,30	23,81	16,85	22,48	25,19	14,29	12,72
Factor ϕ	0,09	0,09	0,13	0,10	0,12	0,11	0,10	0,17	0,31	0,24
Sesgo (compuesto)	0,44	5,11	3,19	6,25	1,30	4,69	7,08	6,13	11,45	-3,31
ECM (compuesto)	14156748	4E+06	816759	975712	136528	114119	173271	90146	18621	25632

Fuente: elaboración propia.

Método de razones de correlaciones: Se estima el total poblacional del año 2011 en el departamento del Chocó (área o dominio de interés) utilizando como insumos los totales poblacionales y como variable sintomática el número de nacimientos en el departamento; esta última variable hace parte del registro de estadísticas vitales, el cual es mantenido por el DANE y publicado anualmente. Este método requiere que se utilice el total

poblacional resultante de los dos últimos censos de población⁴ (1993 y 2005), así mismo que se obtengan los valores de las variables sintomáticas para estos mismos años. Las estimaciones del cambio en proporción del tamaño poblacional en el período intercensal U para el departamento del Chocó (departamento de interés) y del cambio de la proporción de la variable sintomática en Chocó Z_1 son:

$$U_i = \frac{p_{i1}/P_1}{p_{i0}/P_0} = \frac{454.053/42.888.592}{420.631/39.184.456} = 0,986 \text{ y } z_{i1} = \frac{s_{i11}/S_{j1}}{s_{i10}/S_{j0}} = \frac{5.181/719.358}{3.166/720.984} = 1,64$$

Estas mismas variables se calculan para todos los departamentos de

Colombia y se obtiene una ecuación de pronóstico en términos de la variable sintomática:

$$\hat{U}_i = 0,943 + 0,0593z_{i1}$$

La relación existente entre estas dos variables se presenta en el gráfico 1. Para el departamento del Chocó, se pronostica el cambio en la pro-

porción de población del año 2005 al 2011 (cantidad desconocida) en el mismo Chocó, utilizando las estimaciones obtenidas en el modelo anterior y calculando

$$x_{ij} = \frac{s_{ijt}/S_{jt}}{s_{ij1}/S_{j1}} = \frac{5.587/665.499}{5.181/719.358} = 1,1656$$

y la tasa de cambio del número de nacimientos (variable sintomática) entre los años 2005 y 2011:

$$Y^*_i = 0,943 + (0,0593) \times 1,1656 = 1,0119$$

Finalmente la estimación de la población del departamento está dada por:

$$\hat{p}_{it} = 1,0119 \times \left(\frac{454.053}{42.888.592} \right) \times 46.044.601 = 493.264$$

Método de diferencias de correlaciones: Este método, a diferencia del anterior, no implementa el cálculo del cociente del cambio de proporción de las poblaciones en los períodos intercensales y de las variables sintomáticas; en su lugar

³ Dado que la publicación de los saldos migratorios se hace por quinquenios, esta estimación corresponde a un ajuste simple de los saldos netos de migración entre los años 2005 y 2010.

⁴ Para el año 2005 el total poblacional y el número de nacimientos se encuentran disponibles; sin embargo, para el año 1993, año del penúltimo censo, no se encuentran valores publicados del número de nacimientos; por lo tanto, en este ejercicio empírico se utiliza el número de nacimientos ocurridos en el año de 1998 (año más cercano al censo de 1993 en el cual se encuentra disponible la información) y la estimación del total de población realizada por el DANE para ese mismo año a través de proyecciones de población.

calcula la diferencia de los cambios de proporción de la población y de las variables intercensales.

Para el departamento de Chocó se obtienen las siguientes estimaciones:

$$U_i = \frac{p_{i1}}{P_1} - \frac{p_{i0}}{P_0} = -0,0001478 z_{i1} =$$

$$\frac{s_{i11}}{S_{j1}} - \frac{s_{i10}}{S_{j0}} = 0,002811$$

Estas mismas variables se calculan para todos los departamentos de

Colombia y se obtiene una ecuación de pronóstico en términos de la variable sintomática:

$$\hat{U}_i = -0,0533z_{i1}$$

La gráfico 2 presenta la relación entre U y Z.

Para el departamento del Chocó (dominio de interés) se predice el cambio en la proporción de población del año 2005 al 2011 (cantidad

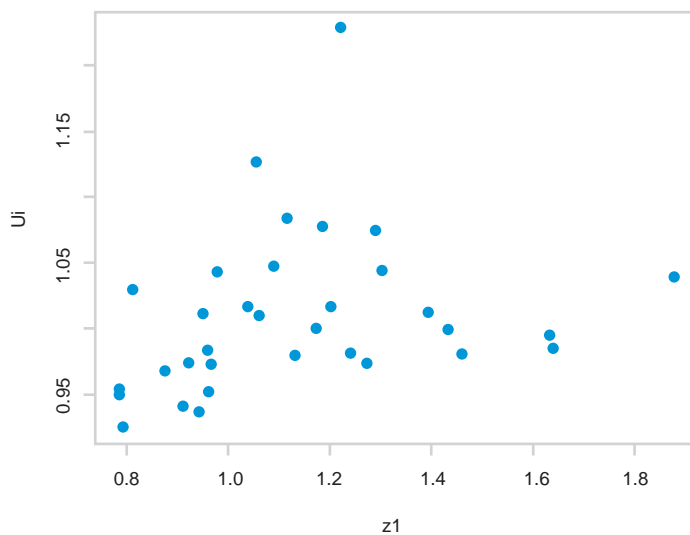
desconocida) utilizando las estimaciones obtenidas en el modelo anterior y calculando $x_{ij} = 0,00119$ la tasa de cambio del número de nacimientos (variable sintomática) entre los años 2005 y 2011:

$$Y^*_i = -0,0000636$$

Finalmente la estimación de la población del departamento está dada por:

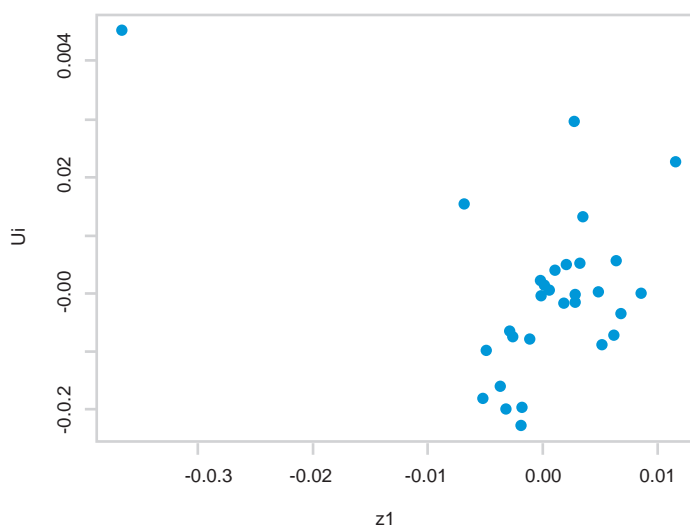
$$\hat{p}_{it} = 490.394$$

Gráfico 1. Diagrama de dispersión de U_i contra Z_1 .



Fuente: elaboración propia.

Gráfico 2. Diagrama de dispersión de U_i contra Z_1 .



Fuente: elaboración propia.

Método de regresión muestral:

Para realizar el método de regresión muestral se utilizan los datos extraídos de la Gran Encuesta Integrada de Hogares (GEIH) para el periodo de 2011 (publicada por el DANE en su página web), la cual proporciona estimaciones para los departamentos del país a excepción de los antiguos territorios nacionales, departamentos que no cubre el operativo de campo y en donde se encuentran localizados menos del 5% de los habitantes del país. Utilizando los factores de expansión asociados al diseño muestral se estima para cada uno de los departamentos la población correspondiente al año 2011. Siguiendo la metodología descrita anteriormente, se elabora un modelo lineal de regresión, el cual permite predecir la población para el año 2011 en términos del número de nacimientos del mismo año para cada uno de estos departamentos:

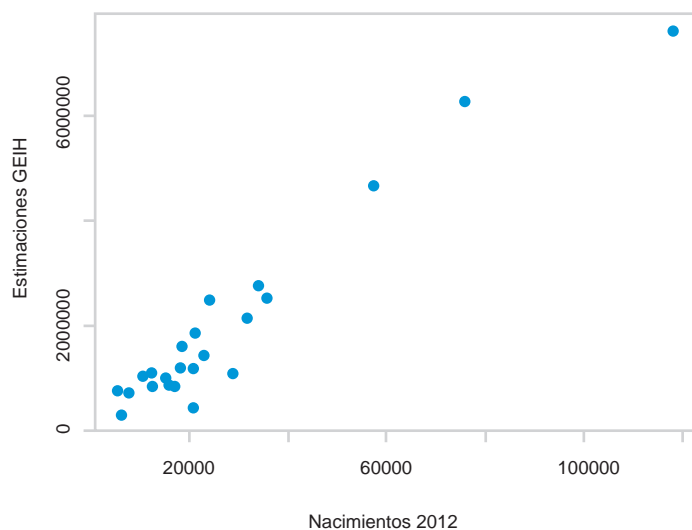
Utilizando los valores de la tabla 5 se ajustó un modelo de regresión—cuya variable dependiente fue la población departamental y su variable independiente fue el número de nacimientos— que permitió la estimación y ajuste del total poblacional para todos los departamentos. En particular, el ajuste que se obtuvo para el departamento del Chocó fue de 399.525 (que resulta de utilizar la ecuación de regresión que se obtuvo con los datos de la tabla 5). La relación

Tabla 5. Estimación de la población por departamento según la GEIH y número de nacimientos en los departamentos según las proyecciones de población.

Departamento	Población estimada GEIH (2011)	Nacimientos (2011)
Antioquia	6272431	75781
Atlántico	2511377	35629
Bogotá, D.C.	7602401	118137
Bolívar	2758392	33854
Boyacá	831654	17217
Caldas	1022907	10556
Caquetá	694128	7537
Cauca	1013583	15183
Cesar	445194	20655
Córdoba	1090520	28774
Cundinamarca	2482754	23878
Chocó	738870	5587
Huila	1162676	20494
La Guajira	849812	12586
Magdalena	1410319	22903
Meta	846722	15855
Nariño	1608545	18479
Norte de Santander	1830953	21252
Quindío	284812	6268
Risaralda	1098922	12279
Santander	2148867	31577
Sucre	946249	15432
Tolima	1179187	18234
Valle del cauca	4660729	57312

Fuente: DANE. Elaboración propia.

Gráfico 3. Diagrama de dispersión de proyecciones de nacimientos a nivel departamental contra la estimación de la población departamental.



Fuente: elaboración propia

existente entre estas variables se presenta en el gráfico 3.

Estimación directa: La estimación del total poblacional en un dominio de interés, en este caso el departamento del Chocó para un año específico (2011), requiere en este método que el total poblacional pueda estimarse por medio de una encuesta. Adicionalmente el método requiere que exista información auxiliar que esté altamente correlacionada con el total poblacional y que la variable auxiliar se encuentre disponible para el dominio de interés, por ejemplo actualizada en un registro administrativo; así mismo, la variable auxiliar deberá encontrarse disponible en la encuesta con la cual se estima el total poblacional. Se pretendía considerar en esta aplicación como variable auxiliar para estimar el total poblacional, el número de nacimientos, información que se encuentra disponible anualmente. Esta variable, sin embargo, tiene el inconveniente de que no se encuentra disponible en algunas de las encuestas de hogares del DANE; por lo tanto, no se puede estimar.

Con el propósito de ilustrar la metodología de estimación directa, se utilizan las proyecciones del número de hombres en el Chocó, la estimación del número de hombres en Chocó según la GEIH y la estimación del total poblacional en Chocó según la GEIH. Utilizando la siguiente expresión se encuentra la estimación de interés:

$$\hat{p}_i = t_{xi} \frac{\hat{t}_{yi}}{\hat{t}_{xi}} = 239.672 \times \frac{738.871}{377.219} = 469.953$$

Donde \hat{t}_{yi} es la estimación de la población total en el departamento del Chocó en el año 2011, \hat{t}_{xi} es la estimación del total de hombres en Chocó en el 2011 y \hat{t}_{xi} es la proyección del número total de hombres en Chocó ese mismo año. Utilizando la expresión anterior, se encuen-

tra que la población en el año 2011 para Chocó fue de 469.953.

Estimación sintética: Se estima la población del año 2011 para el dominio conformado por el departamento del Chocó, utilizando la información resultante de la encuesta de calidad de vida. A través del uso del diseño muestral de la GEIH se estima el total de habitantes de Colombia, \hat{t}_y en 45.492.006. Un total que podría utilizarse como información auxiliar para la estimación de la población es el número de nacimientos en el dominio de interés, que en este caso está conformado por los habitantes del departamento del Chocó; sin embargo, dado que no es posible estimar esta cantidad utilizando la GEIH, se utiliza para ilustrar el método el número de hombres en el departamento del Chocó, información que se encuentra disponible en la GEIH del año 2011 (\hat{t}_x) y además se dispone

$$\hat{p}_{iC} = \varphi_i \hat{p}_{i1} + (1 - \varphi_i) \hat{p}_{i2} = 0,044 \times (469.453) + (1 - 0,044) \times 485.871 = 485.871$$

A diferencia de las metodologías demográficas, sí es posible tener una medición de la precisión de las estimaciones estadísticas obtenidas por los tres métodos estadísticos. Dado que se dispone únicamente de los factores de expansión en la GEIH, se procede a estimar la varianza de los estimadores estadísticos a través del método de remuestreo de Bootstrap (Shao y Tu, 1995), que permite obtener estimaciones de la varianza del estimador y de sus

de la proyección de población del mismo año (t_{xi}). Se utiliza entonces un estimador sintético de razón para llevar a cabo esta estimación:

$$\hat{p}_i = t_{xi} \frac{\hat{t}_y}{\hat{t}_x} = 239.672 \times \frac{45.492.006}{22.440.417} = 485.871$$

Estimación compuesta: Finalmente se construye un indicador compuesto que utiliza la estimación directa y la estimación sintética. El ponderador utilizado para las estimaciones $\hat{\phi}_1$, se estima través de la estimación de la varianza de cada uno de los estimadores, lo anterior debido a que el sesgo de los dos estimadores se asume despreciable y así el error cuadrático medio coincide de manera aproximada con la varianza. La estimación para $\hat{\phi}_1$ es aproximadamente de 0,044.

La estimación de la población en el dominio de interés (Chocó) es calculada finalmente como:

respectivos coeficientes de variación (cve). Para este ejercicio particular, los cve de los tres métodos se muestran en la tabla 6.

Las estimaciones obtenidas a través del método compuesto y la estimación sintética tienen un desempeño significativo mejor que la estimación directa; todas las estimaciones cumplen con los estándares de precisión estadística. Finalmente, la tabla 7 muestra las estimaciones resultantes de cada

método. Nótese que la primera entrada corresponde a la sobreestimación obtenida al sumar los factores de expansión sobre los respondientes en Chocó.

4.1. Discusión

En este artículo se realizó un acercamiento riguroso a un problema que es común en la práctica del investigador involucrado en el sector de las estadísticas oficiales o en estudios sociales. La estimación del total poblacional de un área pequeña es un tema que se encuentra abierto a investigación y es uno de los tópicos de mayor desarrollo en la literatura estadística.

En este artículo se compararon empíricamente algunos métodos tradicionales utilizados en la práctica estadística. Por un lado, los métodos demográficos concernientes a teorías determinísticas y por el otro, algunos métodos estadísticos. Los resultados de las simulaciones muestran que es preferible utilizar estimadores directos que cualquier procedimiento demográfico. Cuando se tiene información auxiliar disponible mediante entidades oficiales es una mejor opción optar por cualquiera de los estimadores que se trataron en la sección 2.2, puesto que, si bien algunos de ellos pueden tener una varianza elevada, conservan la buena propiedad del insesgamiento a medida que los supuestos y las condiciones de regularidad se mantienen.

Se recomienda y enfatiza la utilización de estos métodos en un nivel más profundo de desagregación. Los investigadores asociados a los Institutos Nacionales de Estadística, que tienen acceso al conjunto exhaustivo de datos —el cual contiene la georreferenciación de los respondientes— pueden hacer uso de estas técnicas para la estimación de áreas aún más pequeñas, como lo son municipios, localidades-co-

Tabla 6. Estimación del coeficiente de variación de diferentes estimadores del total poblacional para el departamento del Chocó con los métodos estadísticos.

Método	CVE
Estimación directa	3,27
Estimación sintética	0,655
Estimación compuesta	0,657

Fuente: elaboración propia.

Tabla 7. Estimación de la población del departamento del Chocó con métodos demográficos y estadísticos.

Método	Estimación
Estimación por muestreo GEIH	738870
Tasas de vida	500948
Compuesto	441825
Razón de correlaciones	493264
Diferencia de correlaciones	490394
Regresión muestral	399525
Estimación directa	469953
Estimación sintética	485871
Estimación compuesta	485871

Fuente: elaboración propia.

munas o secciones cartográficas. Es imperativo que las instituciones que proveen estadísticas oficiales a nuestros países establezcan un sistema de estimación

de áreas pequeñas que redunde en la actualización permanente de la información del país. Conocer la realidad a un nivel de desagregación mayor permite que

las autoridades locales tomen mejores decisiones en materia de políticas públicas.

Para futuras investigaciones sería adecuado hacer una comparación empírica en términos de estimadores de máxima pseudo-verosimilitud puesto que incorporan la variación proveniente del diseño muestral junto con la de un modelo de superpoblación. Además, es posible considerar modelos que hacen referencia a modelos especificados para la información de un dominio per se; es decir, totales o medias en cada dominio (modelos Tipo A) y modelos que tengan en cuenta la información individual de los elementos pertenecientes a la población (modelos Tipo B) y, por lo tanto, la información auxiliar esté directamente relacionada con estos individuos, mas no con el dominio.

Bibliografía

Bogue, D. J. (1950). A Technique for Making Extensive Postcensal Estimates. *Journal of the American Statistical Association*, 45, 149-163.

Bogue, D. J. & Duncan, B. D. (1959). A Composite Method of Estimating Post Censal of Small Areas by Age, Sex and colour. *Vital Statistics*. Special Report 47, no. 6, National Office of Vital Statistics, Washington DC.

Brewer, K. R. W. (1999). Cosmetic calibration for unequal probability samples. *Survey Methodology*, 25, 205-212.

Brewer, K. R. W. (2002). *Combined Survey Sampling Inference: Weighing Basu's Elephants*. London: Arnold.

Ericksen, E. P. (1974). A Regression Method for Estimating Pop-

ulations Changes in Local Areas. *Journal of the American Statistical Association*, 69, 867-875.

Gutiérrez, H. A. (2009). *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Bogotá, D. C.: Editorial Universidad Santo Tomás.

O'Hare, W. (1976). Report on Multiple Regression Method for Making Population Estimates. *Demography*, 13, 369-379.

Olayiwola, I. (1986). Use of Percentage Change in Small Area Statistics. *Journal of the Royal Statistical Society*, Series D (The Statistician), 35(5), 531-545.

Rao, J. N. K. (2003). *Small Area Estimation*. New York: Wiley.

Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.

Smith, S. K. & Lewis, B. B. (1980). Some New Techniques for Applying the Housing Unit Method of Local Population Estimates. *Demography*, 17, 323-340.

Schaible, W. A. (1978). Choosing Weights for Composite Estimation for Small Area Statistics. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 741-746.

Schmitt, R. C. & Crosetti, A. H. (1954). Accuracy of the Ratio-Correlation Method of Estimating Post Censal Populations. *Land Economics*, 30, 279-280.



Un análisis del trabajo decente como instrumento conceptual

Claudia Carolina Córdoba Currea

Fuente: www.sxc.hu

Un análisis del trabajo decente como instrumento conceptual¹

Claudia Carolina Córdoba Currea: Economista, magíster en economía de la Universidad Externado de Colombia. Investigadora de los Grupos Transversal (ATT) y Temática Social del DANE. Profesora de la Universidad Externado de Colombia.
Correo electrónico: claudia.cordoba@uexternado.edu.co

Fecha de recepción: 31 de octubre de 2012
Fecha de aceptación: 21 de octubre de 2013

Resumen: el trabajo decente es un objetivo de política económica y social de primera línea para los países firmantes de los Objetivos del Milenio (ODM) y los países miembros de la OECD. En este documento se presenta un balance conceptual de los enfoques de empleos atípicos, trabajos precarios, trabajos informales y empleos de mala calidad, que luego se comparan con el enfoque de trabajo decente. Finalmente se concluye que el trabajo decente permite tener ganancias conceptuales y analíticas para el estudio del mercado de trabajo en comparación con los enfoques ya mencionados.

Palabras claves: trabajo decente, empleos atípicos, trabajos informales, trabajos precarios, calidad del empleo, calidad del trabajo e indicadores de Laeken.

Abstract: nowadays, decent work is a frontline objective of economic and social policies, especially for signatories of the Millennium Development Goals (MDGs) and OECD countries. This paper presents a conceptual balance about atypical employment, precarious employment, informal employments and bad quality employments approaches, which are then compared with the decent work approach. Finally, It concludes that decent work approach offers conceptual and analytical gains for labor market analysis, compared to others approaches.

Key words: decent Work, Atypical Jobs, Informal Jobs, Precarious Jobs, Quality of employment, Quality of work and Laeken indicators.

«Diez personas pierden su empleo y otros cientos se sienten inseguros. Se está desarrollando una sensación de menor control, menor voz, y menor certeza. Tratar estos temas y estas preocupaciones, es fundamental para una participación más amplia en los beneficios de la globalización».

Rodgers, 2007_1:1.

¹ Este artículo corresponde al primer producto de la investigación: «Construcción de un Índice Multidimensional de Trabajo Decente para Latinoamérica», financiada por el DANE de Colombia.

1. Introducción

Se han vuelto comunes los tipos de contrato a término definido, estacional, ocasional y de aprendiz. También se han masificado las relaciones de trabajo triangulares y encubiertas, el falso autónomo (o falso cuentapropismo), los contratos por agencias y la subcontratación; debido principalmente a cambios en la estructura económica global². Así mismo, se han hecho más comunes las condiciones precarias del trabajo: salarios bajos, escasa protección frente al despido, falta de acceso a la protección social compartida con el empleador y mínimas posibilidades de diálogo de los trabajadores ante sus empleadores y el Estado (Rodgers y Rodgers, 1989, 10; De la Garza, 2010, 31).

Lo anterior tiene profundas repercusiones en la sociedad, pues las nuevas modalidades laborales están afectando negativamente la dignidad personal, frustrando las aspiraciones de los individuos y socavando las democracias (Judth, 2010, 137). De igual manera, las nuevas formas de trabajo están desarticulando y desestabilizando las familias, pues están «íntimamente conectadas con el proceso de reproducción social, incluyendo la estructura interna de la familia, las redes familiares y la división social del trabajo» (Moore, 1995, 1).

Un ejemplo de las transformaciones en el mercado de trabajo es la inestabilidad laboral, que se manifiesta en una rotación excesiva entre empleos o en largos periodos de desempleo e inactividad; lo cual supone un desper-

dicio de la capacidad productiva, y una pérdida importante de las competencias laborales que se adquieren en el trabajo (OIT, 2012, 5). Esto se traduce en una menor productividad en el futuro y, por lo tanto, en menos oportunidades para prosperar y en una mayor probabilidad de mantenerse en un círculo vicioso de pobreza.

En consecuencia, a nivel mundial, la evaluación y el seguimiento de las políticas encaminadas a la reducción de la pobreza, y sobre todo de la pobreza extrema, se están focalizando en la consecución del objetivo de trabajo decente propuesto por la OIT, el cual consiste en que todas las personas puedan tener acceso a un trabajo decente. Se entiende trabajo decente como «aquella ocupación productiva justamente remunerada y ejercida en condiciones de libertad, equidad, seguridad y respeto por la dignidad humana» (OIT, 1999). En este sentido «la política social basada en trabajo decente es un instrumento esencial para explicar cómo el crecimiento económico realmente se puede traducir en una reducción de la pobreza» (Anker et al., 2003, 3).

Así, la noción de trabajo decente constituye el pilar de acción del primer objetivo del milenio: «Erradicar la pobreza extrema y el hambre» y para esto «Lograr el empleo pleno y productivo, y el trabajo decente para todos incluidas las mujeres y los jóvenes» (ODM1.2). En la misma línea, la OECD ha exigido que sus países miembros «se comprometan con el tema de trabajo decente... [y que este] deje de ser un *slogan*» (OECD, 2009).

Hay dos metodologías para hacer mediciones del trabajo decente: la primera consiste en la generación de un conjunto amplio de indicadores simples que reflejan la dinámica de cada una de las dimensiones del trabajo decente; y la segunda consiste en diseñar un indicador

sintético, comparable entre países, que resuma el comportamiento de dichas dimensiones.

La primera metodología ha sido la más usada. Sin embargo, al dejar a criterio de los investigadores la escogencia del grupo de indicadores para cada dimensión, al igual que su periodicidad y fuente, el seguimiento a nivel interno ha sido muy difícil y dispendioso y la comparabilidad internacional ha sido casi nula. Por otro lado, con el índice sintético, a pesar de que se gana comparabilidad internacional, se pueden perder particularidades importantes de cada país en el momento de la agregación, y la elección del método o criterio en la asignación de las ponderaciones de cada dimensión puede ser demasiado arbitraria y subjetiva.

Siendo así, pensar en la construcción de un índice sintético de trabajo decente implica tener claridad sobre el valor agregado que tiene este enfoque sobre otras mediciones basadas en enfoques alternativos, ya existentes. En este sentido, el objetivo principal de este documento es entender bien el concepto de trabajo decente y comparar las ganancias conceptuales y analíticas que pueden obtenerse con la adopción del enfoque de trabajo decente, en lugar de los enfoques de los empleos atípicos, precarios, informales o el enfoque de calidad del empleo.

La evaluación que se presenta en este primer documento hace parte de la investigación «La Construcción de un Índice Multidimensional de Trabajo Decente para Latinoamérica»; cuyo objetivo general es construir un Índice Multidimensional de Trabajo Decente para Latinoamérica (IMTD), aprovechando el aprendizaje institucional que ha tenido el DANE con la MESEP y la experiencia adquirida con la construcción de los indicadores de pobreza monetaria y el Índice de Pobreza Multidimensional (IPM).

² Caracterizados por economías más abiertas a los flujos de bienes y de capital; por cambios importantes en las tecnologías de producción (Moore, 1995, 1 y Rodgers, 1989, 19), por aumentos en el desempleo (Rodgers, 1989:19), por el nuevo énfasis en los servicios –para los países desarrollados y la profundización del mismo para los subdesarrollados–, por el incremento del peso específico de los elementos inmateriales en el trabajo, por la tercerización de las actividades de la producción que no implican interacción cara a cara, por la creciente necesidad de trabajadores multitarea, por el incremento en las restricciones de la migración legal de trabajadores (Reygadas, 2009, 4-10) y por la primacía de los mercados financieros sobre los mercados laborales (OIT, 2013-1, 14), han generado nuevas formas de trabajo y de contratación en el mercado laboral a escala mundial.

Este documento se compone de las siguientes partes: la primera es la presente introducción. En la segunda se expondrán, en primer lugar, los puntos de encuentro y/o desencuentro entre los conceptos de empleos atípicos, informales y precarios. Luego se incorporarán al análisis los enfoques de calidad del empleo y calidad del trabajo. En la tercera, se presenta el balance del valor agregado del enfoque de trabajo decente, frente a los otros enfoques ya mencionados; y por último, se concluye.

2. Comparación de enfoques: una aproximación conceptual

Los análisis de los mercados de trabajo se han apartado de la lectura de los indicadores tradicionales (Tasas globales, Tasas de participación, Tasas de ocupación y desempleo) debido a que la atención de los investigadores, desde mediados de la década de los sesenta, se ha volcado hacia el estudio de las causas y las consecuencias de las nuevas formas de arreglos contractuales, la masificación de las relaciones de trabajo no asalariadas y la proliferación de las condiciones precarias del trabajo.

Los niveles o el déficit de los empleos atípicos, los trabajos precarios, los empleos informales, los empleos de calidad y el trabajo decente se establecieron como términos frecuentes en la literatura especializada sobre el mercado laboral. Sin embargo, estos términos en muchas ocasiones son tratados como conceptos equivalentes o como conceptos excluyentes, sin

que quede claro cuál es la población que pertenece a uno solo de estos grupos o si la población pertenece a varios de ellos de forma simultánea. Lo anterior ha dificultado hacer un seguimiento sistemático y comparable a la población de cada grupo. Esta situación ha dificultado los diagnósticos laborales y el seguimiento a las políticas.

2.1. Los empleos atípicos

En los países desarrollados, la discusión en cuanto a las transformaciones del mercado de trabajo, desde mediados de la década de los sesenta, giró alrededor de las formas de trabajo *no estándares* o *atípicas* que estaban creciendo en una proporción importante y que se alejaban cada vez más del trabajo seguro y estable protegido por el Estado de Bienestar (para el caso de Europa occidental).

El *trabajo típico* está asociado a un tipo de trabajo «ideal», que en los países desarrollados en algún momento de su historia llegó a ser mayoritario, caracterizado por ser del sector industrial, a tiempo indefinido, subordinado a un solo patrón y empresa, con relación laboral bilateral –clara diferenciación entre patrón y empleado-, de tiempo completo, con seguridad social y protegido por convenios colectivos y/o asociación sindical (Rodgers y Rodgers, 1989, 10; Pacheco, De la Garza y Reygadas, 2010, 31)³.

Opuesto al trabajo típico se encuentra el trabajo *atípico*, el cual no cumple con una o varias de las características mencionadas anteriormente y, por supuesto, se asocia con una imagen de empleo indeseado (Farné, 2002, 3). El empleo atípico «cubre una multitud de formas de trabajo que se desvían de la forma estándar o típica, en él se encuentran los trabajos temporales, casuales, a tiempo parcial, el trabajo a domicilio, el pluriempleo, el trabajo por cuenta propia, el te-

letrabajo y diversas formas de contratación que disfrazan al trabajo asalariado» (Rodgers y Rodgers, 1989:10-11)⁴.

El *Bureau of Labour Statistics*, BLS, definió el trabajo atípico, con énfasis en el aspecto de la contingencia. Es decir, el trabajo atípico surge de la necesidad de los empleadores de responder de forma eficiente a las posibles fluctuaciones de la demanda. El BLS lo definió como:

«[A] quel trabajo en el cual el trabajador no tiene un contrato de trabajo a largo plazo, ya sea de forma explícita o implícita... [y que hace parte de una] técnica gerencial que se basa en emplear trabajadores cuando solamente había una demanda directa e inmediata por sus servicios... [este incluye] una amplia gama de modalidades laborales entre las que se incluyen el trabajo a tiempo parcial, la contratación a través de empresas de servicios temporales, la subcontratación, el autoempleo, el empleo en el sector servicio y el trabajo a domicilio» (Polivka, 1996, 3-4).

Dada la amplitud de posibilidades en que puede materializarse un trabajo atípico, la OIT distingue las formas de trabajo atípico:

- **Por el tipo de arreglos contractuales:** Contrato a término fijo, contrato a corto plazo, contrato temporal o estacional, trabajo a jornal y empleo ocasional.
- **Por la naturaleza de la relación de trabajo:** relaciones de trabajo triangulares y encubiertas, falso autónomo, contratos de agencias o subcontratistas (OIT, 2011, 7).

Por consiguiente, lo que hace a un empleo «atípico» es que tenga un tipo de arreglo contractual y una relación de trabajo diferente a la que fue en su momento mayoritaria en

⁴ Algunos autores consideran que no necesariamente el empleo debe pertenecer al sector industrial. Si cumple con el resto de condiciones expuestas se puede considerar completamente típico.

³ Todas las traducciones son del autor, a menos que se indique lo contrario.

los países industrializados o la que fue considerada una relación normal en la época de la industrialización en los países en desarrollo. De igual forma, es preciso aclarar que existen diferentes grados de atipicidad, pues no es lo mismo que un trabajo sea catalogado como atípico por no cumplir una de las condiciones expuestas que por más de dos o por todas.

2.2. Los trabajos precarios⁵

A diferencia de los países desarrollados, en los países en vías de desarrollo, los cambios en el mercado de trabajo, se concentraron en la exacerbación de *la precariedad* y *la informalidad*, fenómenos de larga tradición en las economías latinoamericanas (De la Garza, 2010, 19 y Reygadas, 2009, 12).

Según Rodgers (1989, 12), *la precariedad* se puede detectar o identificar teniendo en cuenta cuatro aspectos:

- **Aspecto temporal:** El trabajo carga alta incertidumbre en cuanto a su continuidad y pago oportuno.
- **Aspecto organizacional:** El trabajador tiene escaso control sobre sus condiciones de trabajo en cuanto al salario, al ritmo, la intensidad y la jornada.
- **Aspecto de protección:** El trabajador está débil o nulamente protegido por la Ley laboral ante la discriminación y los tratos inaceptables. Asimismo, el trabajador no tiene acceso a los convenios colectivos ni a la seguridad social.

- **Aspecto económico:** El trabajador recibe ingresos que lo hacen vulnerable a la pobreza y la inseguridad.

Según la OIT:

«El trabajo precario es aquel que se caracteriza por la incertidumbre y la inseguridad...que acarrea en cuanto a la duración del empleo, la presencia de varios posibles empleadores, una relación de trabajo encubierta o ambigua, la imposibilidad de gozar de la protección social y los beneficios que por lo general se asocian con el empleo, un salario bajo y obstáculos considerables tanto legales como prácticos para afiliarse a un sindicato y negociar colectivamente». (OIT, 2011, 5).

Por lo tanto, a diferencia de los empleos atípicos, los trabajos precarios se definen más allá del arreglo contractual o de la naturaleza de la relación de trabajo, pues lo que hace a un trabajo «precario» es el hecho de que le imprima al trabajador una visión de corto plazo por la inestabilidad y una alta vulnerabilidad, debido a que es el trabajador quien asume la total o la mayoría de los riesgos y las responsabilidades en la relación contractual. Así mismo, la precarización se refuerza si desde su posición legal, el trabajador posee poca voz y poder de decisión sobre sus condiciones de trabajo y su remuneración o está sometido a una relación de trabajo asalariada encubierta o ambigua.

Obviamente se espera que, «entre más atípico sea el trabajo, mayor es la probabilidad de que sea precario» (Rodgers y Rodgers, 1989, 11). Dentro del espectro entre atipicidad

y precariedad se pueden encontrar empleos atípicos y no precarios, y trabajos típicos pero precarios.

Para la dupla atípico-no precario, por ejemplo, existen trabajos que son a tiempo definido (atípicos) pero que no acarrearán una alta incertidumbre en la continuidad ni en los pagos y su relación de dependencia asalariada es clara (no precarios). Por ejemplo:⁶

Teresa trabaja en una universidad privada de planta, a tiempo completo pero con contrato fijo anual (atípico). En la Universidad no existe sindicato. Si su rendimiento es bueno (cumple con las publicaciones exigidas y es bien evaluada), con una probabilidad cercana al 100%, su contrato se reanuda en enero, de nuevo por un año. Por su trabajo recibe un buen salario, si lo desea puede hacer parte de los órganos de decisión de sus condiciones laborales y sus pagos siempre son regulares y puntuales (no precarios).

Por el contrario, existen trabajos que pueden ser «típicos» pero que aun así cargan condiciones de precariedad:

Jesús trabaja desde hace 25 años en una de las empresas de mezclas de gases industriales de la Sabana, su contrato es a tiempo indefinido, es miembro del sindicato, es de tiempo completo y su empleo le cubre la seguridad social (típico). Aunque la empresa lleva cerca de 60 años en el país, en los últimos años se convirtió en una filial de una empresa chilena. Con las nuevas condiciones de los chilenos, el sindicato se ha debilitado mucho y los ajustes salariales no han sido buenos; sobre todo para aquellos que no tienen estudios universitarios. En realidad su salario actual no le alcanza para cubrir sus necesidades básicas (precario).

⁵ El término trabajo reúne el conjunto de actividades humanas, remuneradas o no, que producen bienes o servicios en una economía. Es decir, abarca no solo el empleo asalariado, sino también el autoempleo y el trabajo en casa, toda la gama de actividades de la economía informal y las tareas domésticas. El término empleo cubre únicamente al tipo de trabajo que es efectuado a cambio de una remuneración. Por lo tanto, la noción de trabajo transmite una idea más completa de participación en la economía y la comunidad (OIT, 2008; Levaggi, 2006; Féliz y Neffa, 2006 citados en Camacho, Dussán y Guataquí [2012, 18] y Rodgers, 2007, 5).

⁶ Los ejemplos fueron elaborados a partir de entrevistas informales realizadas por la autora. Los nombres y los datos específicos de las personas que brindaron información fueron cambiados, con el fin de cumplir con la reserva estadística.

Las diferentes relaciones *intermedias* entre los empleos atípicos y los trabajos precarios se presentan a continuación en el diagrama 1:

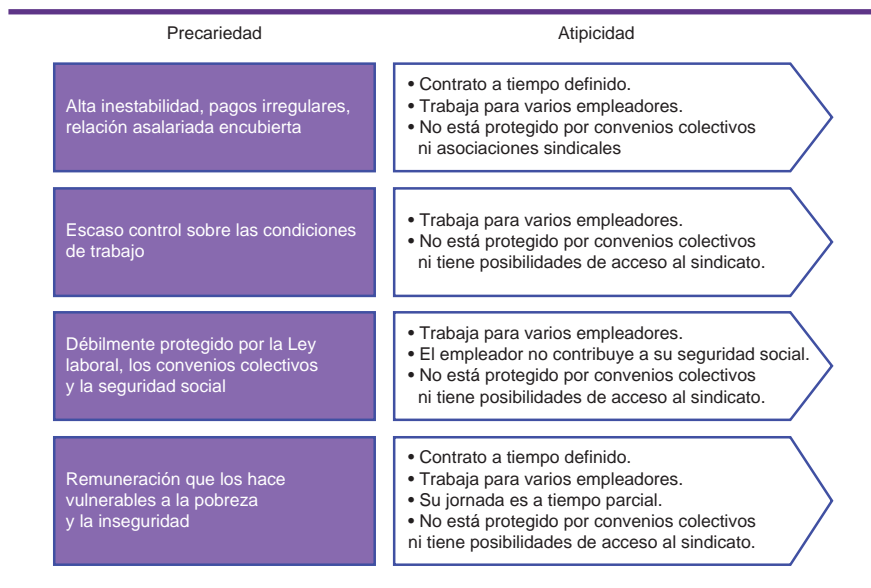
En la primera relación se muestra cómo un empleo puede estar asociado con una alta inestabilidad (precario), si el trabajo no es a tiempo indefinido o no está subordinado a un solo patrón o empresa, o no está protegido por convenios colectivos o por una asociación sindical (atípico). De igual forma, en la cuarta relación se ve cómo

un empleo puede acarrear grandes probabilidades de no generar los ingresos suficientes si no es a tiempo indefinido, si no está subordinado a un solo patrón o empresa, si no es de tiempo completo o si no está protegido por el sindicato. Los casos de Teresa y Jesús no aparecen en el Diagrama 1 pues, estos representan los casos 'extremos'. En el caso de Teresa ambos conceptos están completamente disociados; y en el caso de Jesús operan en conjunto.

salubridad del espacio de trabajo, a las fuertes restricciones que sufren los trabajadores en cuanto a la protección de sus derechos legales y a las limitaciones en el acceso a la información que imposibilitan su movilidad laboral ascendente.

Según Tucker (2002) y Arnold-Bangiovi (2011), el empleo informal se caracteriza por: Bajos e irregulares ingresos, exclusión de la protección social y baja seguridad en el trabajo, bajos niveles de habilidades y productividad que operan en unidades económicas a una baja escala, baja o nula representación colectiva, con condiciones de trabajo inseguras e insalubres y con poco acceso a la información, los mercados, las finanzas y la capacitación formal (Citado en Ferreira, 2012, 14-15).

Diagrama 1. Conexiones entre los trabajos precarios y los atípicos.



Fuente: elaboración propia.

En conclusión, los conceptos de trabajo precario y empleo atípico no son estrictamente equivalentes. Si bien, existe una fuerte asociación entre ellos, no está bien hacerle seguimiento a la población con empleos atípicos para evaluar la precariedad.

2.3. El trabajo informal

La informalidad laboral es una modalidad urbana. No obstante, en este enfoque se le da especial importancia al nivel de escala de la producción, a las limitaciones en el desarrollo de la productividad que hace al trabajador vulnerable a su fácil reemplazamiento, a la

Las dificultades de poner en práctica la definición de informalidad llevaron a que los empleos informales sean estadísticamente identificados de acuerdo con las categorías de «estado-de-empleo» (Ferreira, 2012, 12). Según los lineamientos de la Conferencia Internacional de Estadísticos del Trabajo en 2003, un trabajador es informal si independientemente de que trabaje en el sector formal o informal⁸ o en hogares, cumple con alguna de las siguientes condiciones: (1) es empleador o dueño de su propia empresa del sector informal, (2) es un trabajador familiar auxiliar [sin remuneración], (3) es miembro de cooperativas de productores informales, (4) es trabajador por cuenta propia que produce bienes exclusivamente para el propio uso final de su hogar.

A partir de recomendaciones de la OIT y el grupo DELHI, en Colombia, *el trabajo informal* se mide como⁹: 1) los empleados particulares y los obreros que laboran en establecimientos, negocios o empresas que ocupen hasta cinco personas en todas sus agencias y sucursales, incluyendo al patrono y/o socio; 2) los trabajadores fami-

⁸ El *empleo informal* es diferente al *sector informal*. Este último es definido como un grupo de unidades de producción que forman parte del sector hogares; es decir, empresas que pertenecen a los hogares y que no están constituidas en sociedad. Se diferencian de las sociedades y de las cuasi-sociedades por la organización jurídica de las unidades y el tipo de contabilidad que llevan. Las unidades de producción del sector informal presentan rasgos característicos de las empresas de hogares. El activo fijo y otros valores no pertenecen a la empresa en sí, sino a sus propietarios. En muchos casos es imposible distinguir claramente entre la parte de los gastos asignable a las actividades de producción de la empresa y la que corresponde simplemente a los gastos normales del hogar. Asimismo, ciertos bienes de equipo, como los edificios o los vehículos, pueden utilizarse indistintamente para fines comerciales y para uso propio del hogar (Definición y clasificación del Sistema de Cuentas Nacionales de las Naciones Unidas «Rev. 4»).

⁹ La definición adoptada por el DANE para la medición del empleo informal se remite a la resolución 2 de 1993 de la OIT y las recomendaciones del grupo DELHI (Grupo de expertos sobre estadísticas del sector no estructurado).

liares sin remuneración o trabajadores sin remuneración en empresas o negocios de otros hogares que laboran en establecimientos, negocios o empresas que ocupen hasta cinco personas; 3) los empleados domésticos que laboran en establecimientos, negocios o empresas que ocupen hasta cinco personas; 4) los jornaleros o peones que laboran en establecimientos, negocios o empresas que ocupen hasta cinco personas; 5) los trabajadores por cuenta propia que laboran en establecimientos que ocupen hasta cinco personas, excepto los independientes profesionales; 6) los patrones o empleadores en empresas de cinco trabajadores o menos; 7) se excluyen los obreros o empleados del gobierno (DANE, 2013, 1).

A diferencia de la relación entre los empleos atípicos y precarios, el grado de asociación entre los trabajos precarios y los informales es mucho más alto; incluso, hay autores que los trabajan sin problema como sinónimos (Olmedo, 2006), dado que la precarización se asocia más con las características del empleo informal, según PREALC-OIT¹⁰. Sin embargo, aunque la relación es más estrecha entre ambos conceptos, se presentan casos en que no son estrictamente equivalentes.

Por ejemplo, pueden existir trabajos que son precarios pero que son formales:

Luis es profesional y trabaja para un Ministerio del Gobierno (formal)¹¹, a través de un contrato de prestación de servicios (atípico). Sin embargo, la información y las reglas sobre la continuidad de contrato no son claras, lo que

lo expone a una alta incertidumbre en cuanto a la estabilidad en su empleo; además, por no hacer parte de la planta, no tiene la oportunidad de asociarse sindicalmente o al menos ser parte de los organismos decisores de las condiciones de su trabajo. Además, sus obligaciones en cuanto al horario de trabajo son equivalentes a las de sus colegas asalariados (precario).

En el otro extremo, existen trabajos que son informales mas no precarios:

Alicia trabaja desde hace 14 años para una IPS como aseo, en un municipio de la Sabana de Bogotá. Los empleados del centro de salud primaria, incluido el médico-gerente, suman 5 personas. Su cargo y sus funciones no han cambiado mucho desde que empezó a trabajar en la IPS (informal); posee contrato a tiempo definido con todas las prestaciones de Ley (atípico), y en los 14 años que lleva trabajando en la institución nunca ha tenido incertidumbre ni problemas con la renovación de su contrato, ni con la puntualidad en sus pagos y su sueldo. Además, el ingreso que recibe le alcanza bien para llevar y darle una buena vida a sus dos hijos (no precario).

En conclusión, los trabajadores que laboran en unidades a pequeña escala son más vulnerables a la precariedad, debido a que en las pequeñas unidades de producción, por un lado, los procesos de formación productiva son más escasos, y por otro lado, las actividades a desarrollar son de fácil aprendizaje, situaciones que refuerzan la inestabilidad laboral, la alta rotación, el fácil reemplazamiento y las limitaciones en la movilidad ascendente. Asimismo, a este nivel de escala de la producción, las reglas informales suelen primar sobre las formales,

en cuanto a la resolución de conflictos y el establecimiento de las condiciones del trabajo.

En el diagrama 2¹², se sintetizan las intersecciones y complementos que se pueden presentar entre la población ocupada según la caracterización de su trabajo

El punto A es el que, por tradición, más se ha asociado con el tipo de empleo «ideal o típico». A este se asocian los trabajadores que laboran en medianas o grandes empresas, con contrato a tiempo indefinido, para un solo empleador, con cobertura de seguridad social y con una carrera profesional dentro de la empresa.

En el B están los empleos típicos, formales, pero precarios; representa, por ejemplo, el caso de *Jesús* en la empresa de gases de la Sabana. El punto E representa a los trabajos atípicos, formales pero precarios; en él se podría ubicar a *Luis* con su trabajo en el Ministerio. Las dos anteriores modalidades (B y E) han ganado participación en el mercado de trabajo, y no se les ha prestado la atención que merecen. Estas modalidades de trabajo representan la real precarización de los trabajos ideales del punto A. Y en especial, las más afectadas han sido las nuevas generaciones de trabajadores profesionales.

«Una parte del incremento del empleo no protegido no corresponde a actividades tradicionales del sector informal sino más bien a formas emergentes de empleo asalariado no protegido y empleo semi-dependiente en los sectores “modernos” de la economía. Estas formas van desde el empleo asalariado ilegal hasta arreglos de subcontratación con trabajadores formalmente independientes que en algunos casos se pueden considerar como empleo asalariado disfrazado más que como verdadero

¹⁰ Programa para el empleo de América Latina y el Caribe.

¹¹ Aquí la clasificación de formalidad se basa estrictamente en la medición; a pesar de que para cuestiones de política se utilice el criterio de la afiliación a pensiones para hacerle seguimiento al grupo.

¹² Para la elaboración de este diagrama, se tomó como base la figura «Hypothetical overlaps and distinctions between informality and precariousness», presentada en Ferreira (2012, 29).

trabajo cuenta propia» (Rodgers y Reinecke, 1998, 198).

El punto C es un caso muy común en el que los trabajos cumplen con los estándares de trabajos típicos y no precarios pero que operan a pequeña escala (microempresas), por lo que quedan catalogados en el sector informal. Una modalidad del punto C es el punto F, en el cual se ubican los trabajos atípicos, informales y, aun así, no precarios. No obstante, en los casos en que existe una pobre responsabilidad social empresarial o faltan protocolos formales en las empresas familiares, existe una alta probabilidad de que los trabajos que absorben los puntos C y F se vayan al D; es decir, se precaricen.

En el punto G se encuentran los «peores trabajos», pues son los que están en el círculo vicioso entre la atipicidad, la precariedad y la informalidad; en consecuencia,

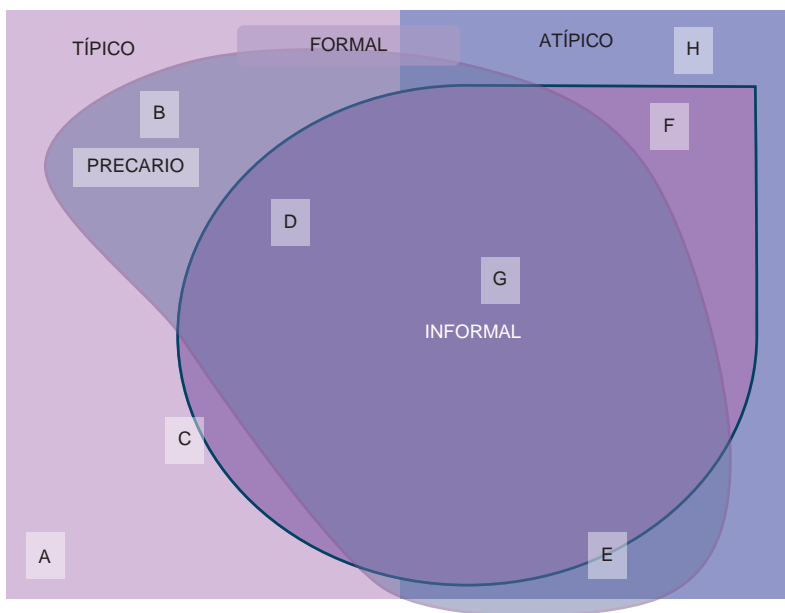
los trabajadores que pertenecen a este grupo son los más vulnerables a caer en el círculo vicioso de la pobreza. Los trabajadores que pertenecen a este grupo no tienen contrato de trabajo, su trabajo no les garantiza el acceso a la seguridad social, sus labores caracterizadas por la baja productividad los hace vulnerables a la inmovilidad laboral y al fácil reemplazamiento, sus ingresos son inestables, etc.

El punto H representa a la nueva «mejor categoría» de trabajos, preferidos hoy por una proporción importante de trabajadores calificados. En esta categoría (la H) se encuentran los trabajos atípicos que se encuentran dentro del empleo formal y no son precarios. Por ejemplo, aquí están los trabajadores con contrato a tiempo definido o prestación de servicios, que no necesariamente pertenecen a un sindicato y/o tienen jornadas de trabajo irregulares

o trabajan desde sus casas (teletrabajo), y no por esto son sujetos de inestabilidad, ni irregularidad en su contratación, ni discriminación y reciben buenas remuneraciones.

En conclusión, las poblaciones con empleos atípicos, precarios e informales sí pueden traslaparse; sin embargo, hay que tener cuidado con los «excesos de generalización», pues no está bien capturar a la población con trabajos precarios a través de los indicadores de informalidad; y ni mucho menos a través de la población que tiene empleos atípicos. Dependiendo de las particularidades del mercado laboral de cada país, los errores en estas generalizaciones pueden variar. A la atipicidad la define su no tipicidad, mientras que a la precariedad la definen la incertidumbre, el riesgo, la desprotección social, la inestabilidad y las relaciones asalariadas encubiertas. Finalmente, la informalidad es definida por el tamaño de la unidad de producción, que está asociada a una baja productividad, escasa formación de nuevas habilidades y las carencias de información que le impiden al trabajador la movilidad laboral hacia arriba.

Diagrama 2. Relación entre los empleos atípicos, precarios e informales.



Fuente: elaboración propia.

- a. Trabajos típicos, formales y no precarios
- b. Trabajos típicos, formales pero precarios
- c. Trabajos típicos, informales mas no precarios
- d. Trabajos típicos, informales y precarios
- e. Empleado atípico, formal pero precario
- f. Empleado atípico e informal pero no precario
- g. Atípico, precario e informal
- h. Atípico pero no formal ni precario

2.4. La calidad del empleo

La imposibilidad de catalogar a los «buenos trabajos» únicamente como los que están en el grupo A y la posible exclusión de población al hacer generalizaciones únicamente tomando como referencia a los trabajos precarios o a los trabajos informales o a los atípicos como los «malos trabajos» han llevado a que la discusión gire hacia el concepto de *calidad del empleo*.

La inmensa heterogeneidad de los trabajadores imposibilita juzgar o clasificar los trabajos como buenos o malos desde posiciones particulares:

«Individuos que tienen diferente historial, tienen a su vez, una

amplia gama de intereses y habilidades. Algunos tienen talentos artísticos y creativos, otros trabajan muy bien con sus manos. Algunos están dotados con un promedio superior de habilidades intelectuales, mientras que otros no. Algunos les gusta trabajar con gente, otros prefieren trabajar solos. Algunos prefieren trabajar al aire libre, otros prefieren el ambiente de oficina. Las condiciones de trabajo estructuradas son preferidas por algunos mientras que otros las rechazan. *Cada individuo refleja una única combinación de intereses y habilidades*» (Rosenthal, 1989, 2, énfasis mío).

En el mismo sentido, Gruenberg Barry (1980, 247) encontró, por ejemplo, que la satisfacción por los factores extrínsecos (sueldos y beneficios, política de la empresa, relaciones con los compañeros de trabajo, ambiente físico y seguridad laboral) es más fuerte en los trabajadores no calificados, semicalificados y/o creyentes. Mientras que la satisfacción por los factores intrínsecos (reconocimiento, independencia laboral, responsabilidad, posibilidades de promoción) es más fuerte en los trabajadores con mayores niveles de educación y mejores habilidades.

«...los cambios sociales han determinado que los valores de estabilidad y el sentido de pertenencia a la empresa en los trabajadores más educados sean reemplazados por otros como la oportunidad de aprender permanentemente a través de una alta rotación en el empleo, o a que la identidad se construya en función de la historia académica y laboral y no de la pertenencia a una empresa en particular» (Valencia, 2003, citado en PGN (2011, 82).

En respuesta a esto, la estrategia de investigación ya no estaba en discutir los puntos de encuentro

y/o desencuentro entre las poblaciones del mercado de trabajo, sino en encontrar una estrategia de valoración adecuada para distinguir los buenos empleos de los malos empleos. Esta tarea fue asumida de forma distinta por los investigadores: Para los economistas, el indicador fundamental que separa los buenos de los malos trabajos es el nivel de remuneración; mientras que, para los abogados, psicólogos y sociólogos del trabajo el nivel de satisfacción del individuo y la seguridad en el trabajo son los factores claves de ponderación. De esta forma, comienzan a aparecer en la literatura, diferentes definiciones sobre los empleos con calidad.

El trabajo perfecto es «un trabajo con variedad de actividades, bajo estrés, con la producción de algo que se pueda ver, con retos permanentes, con reconocimiento del público, con horario flexible, con un alto estatus social, estabilidad y altos salarios» (Rosenthal, 1989, 2-4). Este tipo de empleo podría asociarse a los que se encuentran asociados con el sector H del diagrama 3, de los cuales ya habíamos hablado.

Gittleman y Howell (1995, 14), desde la óptica económica de las teorías de segmentación del mercado de trabajo (SLM), identificaron cinco categorías de valoración para los trabajos: La primera (y a la que más ponderación le dan) son los ingresos y beneficios; la segunda son el tipo de habilidades requeridas; la tercera, las condiciones del trabajo; la cuarta, el estatus del empleo; y la quinta, el marco institucional. Este tipo de aproximaciones a la valoración de la calidad del empleo pone a muchas de las modalidades de trabajos atípicos y algunas de los precarios como deseables y satisfactorias para el trabajador; en especial, para los trabajadores calificados. Muchos trabajadores prefieren aceptar algún grado de precarización en contraprestación de mayor estatus y libertad.

Raúl ha tenido dos grandes pasiones en su vida: la música y la economía. Actualmente trabaja por productos para el Ministerio de Agricultura de Estados Unidos, pero su trabajo lo hace desde Colombia. Su contratación no es regular durante el año, ni tampoco lo son sus pagos (precario). Sin embargo, para Raúl, es un trabajo excelente, pues le permite dedicarse a la música sin perder el estatus y la carrera profesional que para un economista significa trabajar para una firma americana.

Otros autores retoman algunas condiciones del trabajo típico y las entrelazan con la visión no economicista de los empleos de calidad que enfatizan tanto en los aspectos económicos como en los psicológicos y sociológicos. Por ejemplo, Farné (2002, 11) define a un empleo con calidad como aquel que lleva: altos y crecientes salarios, estabilidad laboral y de ingresos, horario de tiempo completo, seguridad social, y posibilidades de formación y ascenso. En la misma línea, Rodgers y Reinecke (1998, 183) adujeron que «Parece razonable considerar que los empleos sin calificación, sin protección social y con ingresos bajos son malos empleos; o que los empleos estables, interesantes, con buenas oportunidades para avanzar en una carrera son buenos empleos». Desde esta perspectiva, los empleos con mayor calidad son los que son los más «típicos» posibles y además ofrecen valores intrínsecos como las posibilidades de formación y ascenso como aproximación a la satisfacción y la motivación del individuo con su empleo.

Infante y Vega-Centeno definen a los empleos de buena calidad en la misma línea de Rodgers, Reinecke y Farné, como:

«[Los que] deberían tener atributos tales como la existencia

de un contrato de trabajo, que garantice la estabilidad tanto del empleo como de las remuneraciones y la provisión de beneficios adicionales en materia de seguridad social (salud y pensiones) y recreación. Igualmente, deberían incluir la regulación de las horas de trabajo, la minimización del factor riesgo (pérdida del trabajo por causas económicas o de enfermedad) y condiciones de trabajo aceptables. También formarían parte de las especificaciones de un buen empleo las características técnicas del mismo (por ejemplo, el hecho de que el trabajo no sea repetitivo), la autonomía de decisión del trabajador dentro de la empresa, el contenido de supervisión del puesto y la posición del mismo dentro de la organización de la producción» (Infante y Vega-Centeno 2001, 181).

Dos conclusiones importantes:

1. La población que goza de un empleo de buena calidad puede traslaparse con las poblaciones que tienen trabajos atípicos, informales e incluso precarios. Por ejemplo, *Alicia*, la aseadora de la IPS, reveló estar satisfecha con su empleo y lo considera de buena calidad, pues es estable y le ha permitido llevar una vida tranquila y familiar (a pesar de ser considerado un empleo atípico e informal). El caso de *Raúl*, como ya se había mencionado, a pesar de la precarización de su empleo, reveló su alta satisfacción con el mismo, incluso al grado de considerarlo perfecto para él.

2. El enfoque de calidad del empleo adiciona los aspectos de la satisfacción, el nivel de autonomía, la complejidad, la motivación personal, la posibilidad de un desarrollo personal-profesional, el estatus, el reconocimiento y el marco institucional como factores decisivos a la hora de clasificar los buenos y los malos empleos.

2.5. Trabajo decente

Juan Somavia, como director entrante de la OIT, en el año 1999 postuló como política principal, para todos países, la consecución del *trabajo decente*¹³. El trabajo decente debe ser aquel que «promueva las oportunidades para que los hombres y las mujeres puedan conseguir un trabajo productivo en condiciones de libertad, equidad, seguridad y dignidad humana».

Actualmente, en la página oficial de la OIT a 2013,

«[E]l trabajo decente resume las aspiraciones de la gente durante su vida laboral. Significa contar con oportunidades de un trabajo que sea productivo y que produzca un ingreso digno, seguridad en el lugar de trabajo y protección social para las familias, mejores perspectivas de desarrollo personal e integración a la sociedad, libertad para que la gente exprese sus opiniones, organización y participación en las decisiones que afectan sus vidas, e igualdad de oportunidad y trato para todas las mujeres y hombres» (OIT, 2013).

A diferencia del enfoque de calidad del empleo, el de trabajo decente amplía el espectro de análisis a todos los tipos de relación laboral, por eso enfatiza en la palabra *trabajo* y no *empleo*. Además, adiciona al análisis, a los desempleados y a la población económicamente inactiva; por ser ellos, los que

más pueden revelar la falta de oportunidades para conseguir un empleo productivo.

De igual forma, complementa las dimensiones de empleo y protección social del enfoque de calidad del empleo con dos objetivos estratégicos adicionales: *la promoción de los derechos fundamentales en el trabajo y el diálogo social*¹⁴. De esta forma, por un lado, además de los derechos básicos de los trabajadores a recibir seguridad social y protección frente a los riesgos profesionales, se adicionan los derechos a la no discriminación, a un sistema justo de resolución de conflictos, al no trabajo infantil y al equilibrio entre la vida familiar y el trabajo. Por otro lado, amplía aún más el alcance del análisis, en comparación con el de calidad del empleo; pues a través de él, se busca hacer un diagnóstico global de la forma en que todos los actores del país (gobierno, empresarios y trabajadores) están articulando sus fuerzas y posiciones, para el funcionamiento del mercado de trabajo. «El trabajo decente es una reivindicación mundial con la que están confrontados los dirigentes políticos y de empresa de todo el mundo» (OIT, 1999, 1).

El concepto de trabajo decente, a diferencia de todos los anteriores enfoques, pone especial énfasis en el concepto de *la seguridad y la suficiencia*. Autores como Bonnet, *et al.* (2003, 2) configuran el concepto de trabajo decente a partir de la seguridad del mercado de trabajo, seguridad del empleo, seguridad profesional, seguridad en el trabajo, seguridad de formación profesional, seguridad de los ingresos y seguridad de representación. Así, desde la perspectiva de trabajo decente se busca, en especial, paliar la incertidumbre y la vulnerabilidad que genera la precarización.

El trabajo decente se enmarca en cuatro grandes dimensiones: 1) la importancia de acceso a empleos

¹³ «Decente es una palabra que no siempre se traduce bien; en otros idiomas el concepto no es exactamente igual y a veces se interpreta "decente" en su sentido estricto de opuesto a lo "indecente", con sugerencias morales. Sin embargo, esa no es su acepción más corriente en inglés, donde el vocablo refleja algo *suficiente* y *deseable*. El trabajo decente puede significar también no tener que esforzarse excesivamente y la posibilidad de un retiro favorable [...] La palabra "decente" también lleva implícitos los estándares de cada sociedad» (Rodgers, 2007, 7-10, énfasis mío).

¹⁴ «La idea matriz de "trabajo decente" es una propuesta integrada de diferentes aspectos, que hacen a la "calidad y cantidad de los empleos", pero también, y de manera fundamental, a las relaciones sociales y en última instancia al tipo de sociedad y de desarrollo al que se aspira» (Espinoza, 2004, 6).

productivos para todos; 2) seguridad en el trabajo, en el ingreso y en el lugar de trabajo; 3) respeto de los derechos fundamentales en el trabajo, incluyendo la ausencia de coacción, discriminación y libertad de asociación; y 4) diálogo social en un proceso democrático (Rodgers, 2007, 6).

El *acceso a empleos productivos para todos*¹⁵ se refiere a que todas las personas que deseen trabajar encuentren un trabajo. Este trabajo debe ser productivo, y debe utilizar al máximo posible sus habilidades y conocimientos. Por su trabajo, las personas deben recibir ingresos dignos y justos por su aporte a la sociedad; además, el trabajo debe asegurarles al trabajador y a su familia una existencia conforme con la dignidad humana, con jornadas justas, con buena seguridad y condiciones de trabajo saludables. Asimismo, debe proporcionarles al trabajador oportunidades de formación y de traslado en pro de ayudarle a su crecimiento profesional.

En materia de *seguridad en el trabajo, en el ingreso y en el lugar de trabajo* se tiene que el trabajo debe brindar una salvaguarda a los trabajadores y sus familias, en los casos en que se presenten contingencias como la enfermedad o los accidentes laborales, la muerte, la vejez, el desempleo y la invalidez. También debe garantizar el empleo en los periodos de retiro por maternidad y asegurar que no se dé por concluida la relación laboral por decisión injusta del empleador o por razones económicas de la empresa. De igual forma, es deseable que, por su trabajo, el trabajador reciba apoyo con los ingresos en especie y con la proporción de medios de recreo y cultura adecuados.

En términos de *respeto de los derechos fundamentales en el trabajo*

(*Derechos Laborales*) se hace énfasis en el marco ético y legal para todos los elementos del trabajo decente. Su objetivo es asegurar que el trabajo se asocie con la dignidad, la equidad y la libertad. El trabajador debe tener libertad de asociación, trato justo, no violento y sin acoso en el trabajo, con cierto grado de autonomía en la ejecución de su labor y un sistema justo de resolución de conflictos. En el mismo sentido, el trabajador no debe ser objeto de discriminación, ni obligado a ejercer labores de forma forzosa ni en su edad infantil. Se considera fundamental que el trabajador tenga derecho a llevar un equilibrio entre la vida familiar y el trabajo.

El *diálogo social en un proceso democrático* hace referencia a que los trabajadores puedan expresar su opinión acerca de sus asuntos profesionales y capacidad de intervención en la fijación de sus condiciones de trabajo. Esto implica que existan espacios democráticos para presentar sus puntos de vista, defender sus intereses y negociar con las autoridades del gobierno y los empleadores. Esta dimensión también implica una permanente cooperación entre los hacedores de política, empleadores y trabajadores, con el fin de mejorar continuamente la eficiencia en la producción y la colaboración en las medidas sociales y económicas.

En conclusión, el concepto de trabajo decente al igual que el concepto de calidad del empleo no centra su análisis en la cuantificación y seguimiento de una población particular, a diferencia de los enfoques de atípicos, precarios e informales. Por consiguiente, la población que pueda llegar a ser calificada con empleo decente también se traslapa con los anteriores enfoques; sin embargo, por el enfoque de las dimensiones, se puede inferir que los empleos atípicos y los informales tienen una más alta proba-

bilidad de llegar a ser catalogados como decentes, a diferencia de los precarios, pues estos vulneran de entrada los requerimientos del enfoque de derechos.

A diferencia del enfoque de calidad, el trabajo decente busca dar una calificación general a todos los actores y espacios del mercado laboral, no únicamente al puesto de trabajo. Así mismo, adiciona el enfoque de derechos y el diálogo social en democracia como articuladores fundamentales para valorar las posibles sinergias entre hacedores de política, empresarios y trabajadores.

2.6. Trabajo decente y calidad del trabajo

La Unión Europea en diciembre de 2001, en Laeken, oficializó los indicadores de *Quality of Work Employment* (QWE), con los cuales se integraron al análisis de la calidad del empleo y los objetivos del trabajo decente, ampliando el concepto de la calidad del empleo hacia la *calidad del trabajo*. Con ellos se le da un énfasis especial a la dimensión de «satisfacción en el trabajo». Los indicadores de Laeken en el marco del QWE se agruparon en diez dimensiones (European Commission, 2001; Peñas, 2009):

- 1. Las cualidades intrínsecas del trabajo.** Los trabajos deben ser intrínsecamente satisfactorios y compatibles con las capacidades y habilidades del trabajador, y proporcionar niveles adecuados de ingresos.
- 2. Las habilidades, el aprendizaje permanente y el desarrollo profesional.** El trabajador debe poder ser capaz de desarrollar sus capacidades potenciales al máximo, mediante el apoyo apropiado para el aprendizaje permanente.
- 3. Igualdad de género.** El mercado de trabajo debe ofrecer

¹⁵Esta sección se basa en los trabajos de Anker et al. (2003), Rodgers (2007_1), De justicia (2008), Gamero (2010), PGN (2011), Camacho, Dussán y Gualaqui (2012) y OIT (2013).

igualdad de oportunidades para hombres y mujeres, en cuanto al acceso a trabajos de calidad equivalentes y a las oportunidades de desarrollo profesional.

4. Trabajo saludable y seguro. Las condiciones de trabajo deben ser seguras, saludables y de apoyo, tanto en términos psicológicos como físicos.

5. Flexibilidad y seguridad. Debe existir un equilibrio apropiado entre la flexibilidad y la seguridad en el trabajo, con el fin de fomentar la movilidad laboral. Esto requiere brindar un apoyo apropiado para las personas que pierden su empleo o están buscando otras alternativas, así como el fomento de la plena utilización de las capacidades y las opciones de carrera flexibles a través de un apoyo adecuado a la movilidad profesional y geográfica.

6. Inclusión y acceso al mercado de trabajo. Las oportunidades de acceso e inclusión al mercado de trabajo deben incrementarse, en especial para aquellos que lo hacen por primera vez o que están retornando al mercado después de periodos de desempleo o inactividad; de modo que se les permita su permanencia en el mercado laboral.

7. Organización del trabajo y conciliación de la vida familiar. Deben existir acuerdos adecuados, en especial, en cuanto al tiempo de trabajo, y servicios de apoyo, para que los trabajadores tengan un equilibrio entre la vida laboral y la vida fuera del trabajo.

8. Diálogo social y participación de los trabajadores. Todos los trabajadores deben ser informados y poder tener participación en el desarrollo de sus empresas y su vida laboral.

9. Diversidad y no discriminación. Todos los trabajadores deben ser tratados por igual y sin discriminación por motivos de edad, discapacidad, origen étnico, religión u orientación sexual.

10. Ejecución del trabajo en general. Los trabajos deben ser productivos, de modo que se garanticen altos niveles de vida en todas las regiones.

Como puede apreciarse en el diagrama 3, las diez dimensiones de calidad del trabajo están contenidas en las cuatro dimensiones del enfoque de trabajo decente. Por lo tanto, el trabajo decente puede considerarse un enfoque más general que el de calidad del trabajo; no obstante, en el momento de pensar en indicadores sintéticos para ambos enfoques, el de trabajo decente deberá tener una mayor ponderación en las dimensiones de respeto a los derechos en el trabajo y el diálogo social, mientras que el de calidad del trabajo lo tendrá en la dimensión de la satisfacción intrínseca.

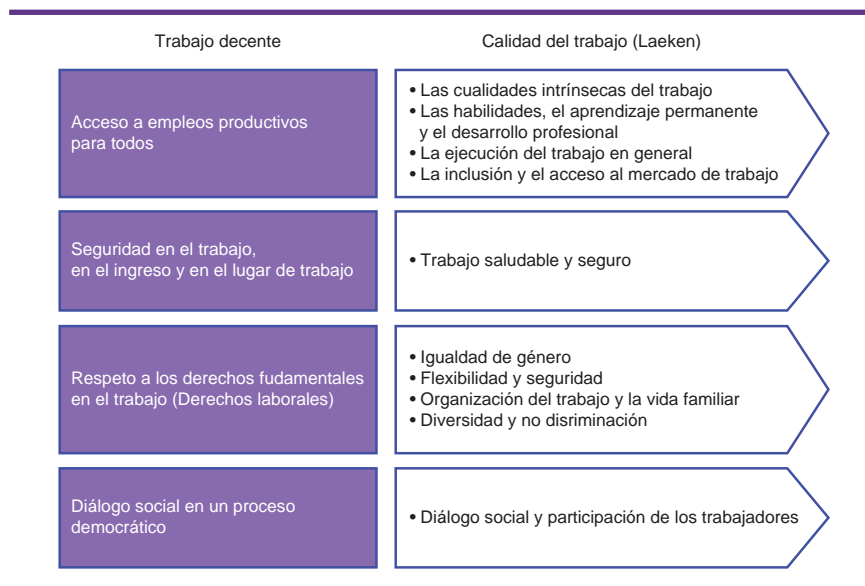
3. El valor agregado del enfoque de trabajo decente

El enfoque de trabajo decente permite tener ganancias conceptuales y analíticas para el estudio del mercado de trabajo.

A través del enfoque de trabajo decente se le puede hacer seguimiento a los empleos atípicos, sin el juicio de valor de que sean malos empleos *per se* (diagrama 4); pues como ya se señaló arriba, hay tipos de contratos que a pesar de que se alejan de la forma tradicional o típica, son deseables y convenientes tanto para los empleadores como para los trabajadores. La decencia se define a partir de la suficiencia y no por lo tradicional. En otras palabras, el enfoque de trabajo decente abre la posibilidad de darle una valoración positiva a las nuevas formas de arreglo contractual y a las relaciones de trabajo modernas, en vista de que, según el contexto, algunas de ellas han mostrado ser buenas para todos los actores.

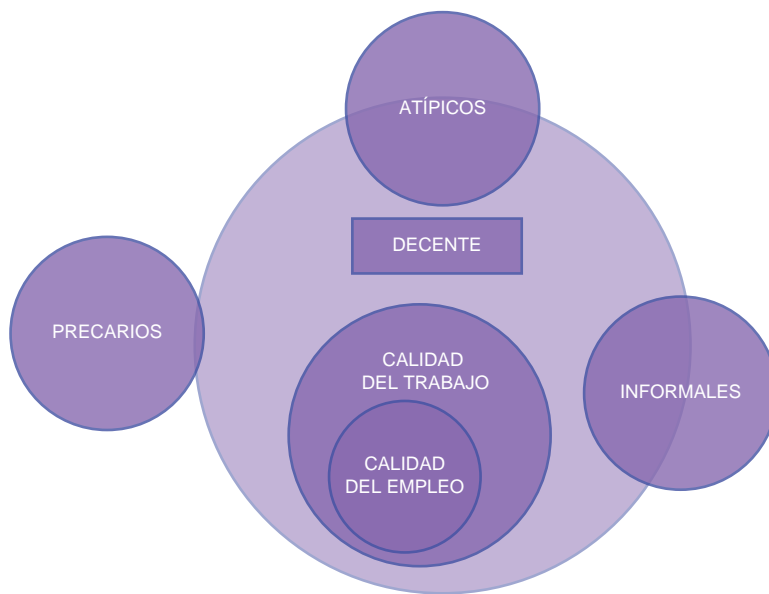
En el mismo sentido, el enfoque de trabajo decente permite hacerle se-

Diagrama 3. Trabajo decente y calidad del trabajo (Laeken).



Fuente: elaboración propia.

Diagrama 4. El trabajo decente frente a los otros enfoques.



Fuente: elaboración propia.

guimiento a los trabajos precarios, pues salvo excepciones subjetivas como el caso de Raúl, trabajos precarios no pueden ser decentes, debido a que el concepto de trabajo decente descansa en el criterio de la seguridad en el ingreso, en el empleo, en el lugar de trabajo y en los derechos laborales en general; todos aspectos carentes en los trabajos precarios. Por esta razón, en el diagrama 4 vemos que la intersección entre los trabajos precarios y decentes es muy pequeña.

Como el trabajo decente no tiene en cuenta la escala de producción, el enfoque permite hacer una valoración de los trabajos formales que violan los criterios de suficiencia y

seguridad. Permite, además, identificar los problemas de información, oportunidades, discriminación e incompatibilidad con las actividades fuera del trabajo, de las que usualmente son objeto una buena parte de los trabajadores informales que no alcanzan un trabajo decente.

Si bien los enfoques de calidad del empleo, calidad del trabajo y empleo decente son tratados por un número importante de investigadores como si fueran equivalentes—debido a que todos incluyen los aspectos del nivel de ingresos y beneficios, las condiciones del empleo, la satisfacción del trabajador con su empleo, el nivel de autonomía, la complejidad, la motivación personal,

la posibilidad de un desarrollo personal-profesional y el marco institucional como factores decisivos en la valoración de los buenos y los malos trabajos—, el enfoque de trabajo decente incluye a los de calidad, pero no se reduce a ellos.

Lo anterior debido a que el enfoque de trabajo decente pone especial énfasis en toda la población del mercado laboral, trabajen o no trabajen, estén activos en el mercado o inactivos. También acentúa el aspecto democrático; es decir, en la voz que tienen todos los actores (Gobierno, empleadores y trabajadores) para exigir y negociar sus derechos. En otras palabras, como se muestra en el diagrama 4, el enfoque de trabajo decente contiene al enfoque de la calidad del trabajo y a su vez este contiene a los enfoques de calidad del empleo.

Finalmente, el enfoque más similar al de trabajo decente es el de calidad del trabajo. Todas las dimensiones del segundo se encuentran en el primero. Según la disponibilidad de datos confiables en cada país se puede partir desde un enfoque de trabajo decente y perfectamente terminar en uno de calidad del trabajo. En especial, porque las estadísticas sobre los derechos laborales (acoso, no discriminación, trabajos forzados) y las de diálogo social, en su mayoría, no hacen parte de las encuestas de hogares, lo que puede determinar su exclusión de los índices sintéticos.

Conclusiones

Todo empleo que no sea a tiempo indefinido, subordinado a un solo patrón y empresa, con una clara diferenciación entre patrón y empleado, de tiempo completo, con seguridad social y protegido por convenios colectivos y/o asociación sindical, se cataloga como atípico.

No obstante, los trabajos atípicos (a tiempo definido, por prestación de servicios, por jornal, a domicilio, por cuenta propia o contratados por agencias o cooperativas etc.) no deben ser *per se* caracterizados por la inestabilidad, la vulnerabilidad, la desprotección social o por las relaciones de trabajo encubiertas o ambiguas. En otras palabras, los trabajos atípicos no tienen por qué ser, por definición, precarios.

Por otro lado, no todos los trabajos informales son precarios. Si bien existe una fuerte conexión entre ellos –debido a que cuando se trabaja a muy pequeña escala, los trabajadores son de fácil reemplazo, vulnerables a recibir remuneraciones precarias y a la desprotección social–, asociar la precarización únicamente con los trabajos informales lleva a una subvaloración de los trabajos en las medianas y grandes empresas que han deteriorado las condiciones de seguridad para los trabajadores o que llevan al trabajador a un falso cuentapropismo.

Lo anterior también evidencia que no existen preguntas específicas

en las encuestas laborales para hacerle seguimiento a la precarización, a pesar de que la población que tiene un trabajo precario son los que, con mayor probabilidad, pueden caer en condiciones de pobreza. Por esta razón, se incurre en generalizaciones indebidas, al catalogar a los trabajadores atípicos o a los informales como precarios.

Como alternativa, se cuenta con el enfoque de la calidad del empleo, el cual le suma a la valoración de los buenos trabajos, los aspectos de la satisfacción, la autonomía, el balance entre el desarrollo profesional y personal, el reconocimiento y el marco institucional. Sin embargo, este enfoque no ha sido adoptado de forma oficial por los departamentos de estadística, debido a que no existe un consenso en cuanto a las variables que se deben tener en cuenta para la medición de los factores adicionales ya mencionados. Así mismo, no hay un acuerdo sobre las ponderaciones que se le deben dar a los mismos. Tampoco es claro si debe existir un criterio unilateral de calidad del empleo para todas las ramas de la economía.

Los enfoques que van más allá de la calidad del empleo; es decir, los de calidad del trabajo y trabajo decente, incluyen en el espectro de análisis a todos los tipos de relación laboral, a toda la población en edad de trabajar y a un mayor número

de derechos tanto para los trabajadores como para los empleadores. Con el enfoque de calidad del trabajo se le otorga especial ponderación al nivel de satisfacción que revele el trabajador con su ocupación; mientras que con el enfoque de trabajo decente se busca darle mayor ponderación a las dimensiones que cubre toda la gama de derechos y al diálogo social.

Al ser la suficiencia y la seguridad los elementos que permiten distinguir los trabajos decentes y los no decentes, se abre la posibilidad de que trabajadores con empleos atípicos, trabajos informales, y en menor medida los trabajadores con empleos precarios, entren al grupo de los que tienen un trabajo decente.

Finalmente, como todas las dimensiones de la calidad del empleo y las de la calidad del trabajo quedan incluidas en las cuatro categorías del trabajo decente, se considera que el enfoque de trabajo decente es la categoría más amplia de análisis. No obstante, los desafíos para su medición son grandes, pues hereda los problemas de escogencia de variables y ponderaciones del enfoque de la calidad del empleo y se enfrenta a la ausencia de información en las categorías que precisamente lo hacen diferente a los enfoques de calidad.

Bibliografía

- Alkire, S. & Foster, J. (2007). Counting and Multidimensional Poverty Measurement. OPHI Working Paper 7.
- (2011). Counting and Multidimensional Poverty Measurement. *Journal of Public Economics*. P. 476-487.
- American Sowunmi F. Akinyosoye V., Okoruwa, Omonona B. (2012). The Landscape of poverty In Nigeria. A spatial analysis using senatorial Districts Level-Data. *American Journal of Economics*. Vol. 2.
- Angulo, R; Díaz, Y. & Pardo, R. (2011). Índice de Pobreza Multidimensional para Colombia (IPM- Colombia) 1997-2010. Archivos de Economía. Departamento Nacional de Planeación. Documento 382. Noviembre.
- Anselin, L. (1992). Spatial data analysis with GIS: an introduction to application in the social sciences. National Center for Geographic Information and Analysis. University of California. URL: <http://repository.topoi.org/BACP-BIB/Anselin1992/Anselin1992.pdf>
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis* (Vol. 413). Essex: Longman Scientific & Technical.
- Bonet Morón J. (2007). ¿Por qué es pobre el Chocó?, Documentos de Trabajo sobre Economía Regional, Banco de la República.
- Buzai, G. D. (2012). Geografía y Sistemas de Información Geográfica: Evolución teórico metodológica hacia campos emergentes. *Revista Geográfica de América Central*, 2(48E), 15-67.
- Buzai, G. D., & Baxendale, C. A. (2006). *Análisis socioespacial con sistemas de información geográfica*. Buenos Aires.
- Cepal-Unicef (2012). *Metodología para la medición de la autocorrelación espacial y la heterogeneidad espacial de la pobreza infantil en América Latina y el Caribe*. Recuperado de <http://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Libro-pobreza-infantil-America-Latina-2010-Anexo-III.pdf>.
- Chasco Yrigoyen, C. (2001). *Análisis Exploratorio de Datos Espaciales al servicio del Geomarketing*. Instituto LR Klein, Universidad Autónoma de Madrid. III Seminario sobre Nuevas Tecnologías en la Investigación, el Marketing y la Comunicación. e-Participación
- (2004). *Modelos de heterogeneidad espacial*. EconWPA.
- (2010). *Métodos gráficos del análisis exploratorio de datos espaciales*. Instituto LR Klein, Departamento de Economía Aplicada, Universidad Autónoma de Madrid. URL: <http://www.asepelt.org/ficheros/File/Anales/2003%20-20Almeria/asepeltPDF/93>.
- Estrada, L. & Moreno, S. L. (2013). Análisis espacial de la pobreza multidimensional en Colombia a partir del Censo de Población de 2005. Documentos de Trabajo Candane. URL: http://www.dane.gov.co/candane/index.php?option=com_content&task=section&id=18
- Fernández, M., Hernández, C., Ibáñez, A. M., & Jaramillo, C. (2009). *Dinámicas departamentales de pobreza en Colombia 1993-2005*. Documento de Trabajo núm. 33. Programa Dinámicas Territoriales Rurales. Rimisp, Santiago, Chile.
- Fujita, M., & Krugman, P. (2004). La nueva geografía económica: pasado, presente y futuro. *Investigaciones regionales*, 4(1), 177-206.
- Galvis, L. A., & Meisel, A. (2010). *Persistencia de las desigualdades regionales en Colombia: Un análisis espacial*. Documento de trabajo sobre economía regional, (120). Banco de la República.
- Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24(3), 189-206.
- Glasmeyer, A. K. (2002). One nation, pulling apart: the basis of persistent poverty in the USA. *Progress in Human Geography*, 26(2), 155-173.
- Goodchild, M. F. (1986). *Spatial autocorrelation*. Norwich: Geo Books.
- Goodchild, M. F., & Haining, R. P. (2004). GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science*, 83(1), 363-385.
- Henninger, N., & Snel, M. (2002). *Where are the poor?: Experiences with the development and use of poverty maps*. WRI; PNUD/GRID.
- Núñez Méndez, J., Ramírez J., Juan Carlos. (2002). *Determinantes de la pobreza en Colombia: años recientes*. Bogotá: Naciones Unidas, CEPAL.
- Núñez, J., Ramírez, J. C., & Cuesta, L. (2005). *Determinantes de la*

pobreza en Colombia, 1996-2004. Documento CEDE, 60.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, 27(4), 286-306.

Pérez, G. J. (2005). *Dimensión espacial de la pobreza en Colombia*. Documentos de trabajo sobre economía regional (54), Banco de la República.

Progress in Human Geography, 26, 155-173.

Sánchez Peña L. (2006). Métodos para el análisis espacial. Una aplica-

ción al estudio de la geografía de la Pobreza. Universidad de Wisconsin Madison. III Congreso de la asociación latinoamericana de Población.

Sánchez, F., & Nuñez, J. (2000). La geografía y el desarrollo económico en Colombia: una aproximación municipal. *Desarrollo y sociedad*, 46, 43-98.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46, 234-240.

Unwin (1975). An Introduction to trend analysis. Norwich: Geo Abstracts.

Unwin, D. J. (1975). An introduction to trend surface analysis. Geo Abstracts Limited.

Vargas Mesa, O. M. (2012). Aproximación espacial al estudio de la pobreza en Colombia: un estudio aplicado con información del año 2005 [recurso electrónico].

Vista, B. M., & Murayama, Y. (2011). Spatial Determinants of Poverty Using GIS-Based Mapping. In *Spatial Analysis and Modeling in Geographical Transformation Process* (pp. 275-296). Springer Netherlands.



Análisis espacial de la pobreza multidimensional en Colombia a partir del censo de población de 2005

**Laura Estrada Arbeláez
Sandra Liliana Moreno Mayorga**

Fuente: www.sxc.hu

Análisis espacial de la pobreza multidimensional en Colombia a partir del censo de población de 2005

Laura Estrada Arbeláez: Economista, magíster en Estudios Latinoamericanos con énfasis en Sociología del Instituto de Altos Estudios para América Latina, Universidad Sorbona Nueva, París 3, Francia. Profesional de la Dirección de Metodología y Producción Estadística del DANE.
Correo electrónico: lestradaa@dane.gov.co

Sandra Liliana Moreno Mayorga: Ingeniera catastral y geodesta, candidata a magíster Sistemas de Información Geográfica en la Universidad de Salzburgo, Austria. Profesional en Investigación y Desarrollo de la Dirección de Geoestadística del DANE.
Correo electrónico: slmoreno@dane.gov.co

Fecha de recepción: 15 de agosto de 2013
Fecha de aceptación: 30 de octubre de 2013

Resumen: en este documento se explora la dimensión espacial de la pobreza multidimensional en Colombia a nivel municipal, usando el Censo de Población de 2005 y el software de Sistemas de Información Geográfica (SIG) ArcGIS 10.1 de ESRI. La metodología incluye elaboración de mapas de pobreza, análisis exploratorio de datos espaciales, análisis de tendencia, y pruebas de autocorrelación espacial global y local (índice de Moran y Anselin local de Moran). Se encuentra que la pobreza multidimensional presenta autocorrelación espacial positiva; es decir, los municipios pobres tienden a estar rodeados de municipios pobres, y viceversa. Se identifican polos y corredores de bienestar en la región andina; conglomerados de pobreza al sur del país y en las costas Pacífica y Caribe, que se caracterizan por ser discontinuos y exhibir una marcada diferencia entre los niveles municipal total, área urbana (cabecera) y área rural (resto); y atípicos espaciales, municipios con bajo nivel de pobreza rodeados de municipios con alto nivel de pobreza y viceversa.

Palabras clave: análisis espacial, pobreza multidimensional, autocorrelación espacial.

Abstract: this document explores the spatial aspect of multidimensional poverty in Colombia on a municipal level using the 2005 Population Census and ESRI's ArcGIS 10.1 Geographic Information Systems (GIS) software. Methodology includes poverty mapping using GIS-based, exploratory spatial data analysis, trend analysis, and global and local autocorrelation tests (Moran's I and Anselin Local Moran's I). The findings suggest that multidimensional poverty demonstrates positive spatial autocorrelation, meaning poorer municipalities tend to cluster around each other, and vice versa. Hubs and paths of prosperity can be found in the Andean Region, while poverty clusters can be found in the south of the country and along the Pacific and Caribbean coasts. These clusters are characterized by being discontinuous and showing a notable difference in poverty levels between urban (cabecera) and rural (resto) areas. Spatial outliers, municipalities with low levels of poverty relative to surrounding municipalities and vice versa, are identified.

Key words: spatial analysis, multidimensional poverty, spatial autocorrelation.

1. Introducción¹

Desde hace varias décadas, la reducción de la pobreza hace parte de la agenda de trabajo de diferentes organismos internacionales, tales como el Banco Mundial, el Banco Interamericano de Desarrollo (BID), la Organización de las Naciones Unidas con su programa para el Desarrollo Humano (PNUD), de agencias de cooperación internacional de gobiernos como el de Estados Unidos, Alemania, España, etc. y de los propios gobiernos nacionales.

Esto ha generado gran interés por explicar la existencia, los determinantes y la persistencia de la pobreza a través de variables como los ingresos gubernamentales, el crecimiento de la población, la migración, la educación, la composición del hogar y la condición laboral. Más allá de los determinantes socioeconómicos, la pobreza tiene un componente espacial (Glasmeyer, 2002) que puede ser explorado a través de los mapas de pobreza («poverty mapping») y el uso de los Sistemas de Información Geográfica (SIG); herramientas que permiten integrar características socioeconómicas y biofísicas al análisis de la pobreza (Heninger y Snel, 2002; Vista y Murayama, 2011).

De esta manera, la motivación principal para la realización de este documento es la vinculación del componente espacial a la comprensión de la pobreza multidimensional en Colombia, usando el índice de pobreza multidimensional (Angulo *et al.*, 2011), calculado a nivel municipal

pal a partir del Censo de Población de 2005 y, usando los Sistemas de Información Geográfica (SIG). Las preguntas que guían esta investigación son: ¿Existe un componente espacial de la pobreza multidimensional en Colombia? ¿Se distribuye esta de manera aleatoria en el territorio o presenta algún patrón espacial? ¿Hay conglomerados de municipios de mayores o menores niveles de pobreza? ¿En dónde se encuentran estos conglomerados? ¿Cómo difieren los resultados para los niveles municipal total, cabecera y resto? ¿Puede el análisis espacial de la pobreza ayudarnos a entender algunas dinámicas territoriales?

Para responder dichas preguntas se realiza un análisis de la distribución espacial de los datos de pobreza; se identifican los conglomerados críticos de pobreza; y se examinan las variaciones de los niveles de pobreza en el territorio para las entidades geográficas total municipal, cabecera² y resto³. Se espera encontrar que la pobreza no se distribuye aleatoriamente en el espacio, y que existen conglomerados de pobreza ubicados principalmente en regiones como las costas Pacífica y Caribe, Amazonía y Orinoquía (Pérez, 2005; Galvis y Meisel, 2010; Vargas, 2012). Adicionalmente, se espera encontrar que los patrones espaciales de la pobreza multidimensional a nivel municipal total, cabecera y resto difieren entre sí, y que a su vez tienen influencia sobre los conglomerados de pobreza. Finalmente se busca identificar y localizar los polos y corredores de bienestar en el territorio colombiano.

El documento se compone de cuatro partes. En la primera sección se presenta la revisión de literatura sobre el análisis espacial de la pobreza en el caso de Colombia. En la segunda sección se presenta la metodología para el cálculo del IPM y el análisis espacial. En la tercera sección se presentan los

resultados; y en la cuarta, las conclusiones del estudio.

2. Revisión de literatura

El estudio de la pobreza ha sido ampliamente abordado desde la economía y las políticas públicas. Para explicar la existencia, los determinantes y la persistencia de la pobreza, y la desigualdad, a nivel nacional y regional en Colombia, se han usado tanto estadísticas descriptivas como técnicas econométricas (Núñez y Ramírez, 2002; Núñez, Ramírez y Cuesta, 2005; Fernández *et al.*, 2009; Bonet, 2007).

En estos las variables más recurrentes para el estudio de la pobreza son los impuestos de industria y comercio, el crecimiento de la población, la migración, la educación (la cobertura, el nivel educativo de la fuerza de trabajo, los años promedio de educación), la composición del hogar (el jefe de hogar, el tamaño) y el mercado laboral (el desempleo, formalidad en el empleo).

La integración del componente espacial al análisis de temáticas sociales, ambientales y económicas ha sido liderada principalmente por desarrollos teóricos tales como la «Nueva Geografía Económica» (Fujita y Krugman, 2004) y «La Ciencia Social Integrada Espacialmente» del Center for Spatial Integrated Social Sciences (CSISS). En estos enfoques se destaca la concepción del espacio no como un elemento contenedor sino como un elemento explicativo (Buzai, 2012).

Inspirados en estos enfoques, los modelos econométricos incluyen variables de la geografía física (temperatura, relieve, calidad del suelo, etc.) y algunas medidas de distancia (tiempo, kilómetros) a los principales centros económicos Bogotá, Medellín, Cali y Barranquilla, y

¹ Los autores agradecen a Álvaro Murillo por su colaboración y orientación brindada en la realización del presente artículo, a Carlos Durán de la Dirección de Geoestadística por la elaboración de los mapas que se presentan en el documento, y a Néstor Muñoz del Centro Andino de Altos Estudios (Candane) por sus valiosos comentarios al documento.

² Según el DANE, la cabecera municipal corresponde al área más densamente poblada del municipio y donde funciona la sede de la Alcaldía Municipal. Su área geográfica está definida por un perímetro urbano, cuyos límites se establecen por «acuerdos» del Concejo Municipal. Esta entidad geográfica se asocia a la zona urbana del municipio.

³ Según el DANE, el resto del municipio corresponde al área que está por fuera del perímetro urbano de la cabecera municipal. Puede estar constituido por centros poblados y población dispersa. Esta entidad geográfica se asocia a la zona rural del municipio.

a los puertos de Barranquilla y Buenaventura, buscando comprender el impacto que tienen dichas variables en la riqueza o bienestar de una población a nivel municipal o departamental (Sánchez y Núñez, 2000).

Sin embargo, el uso de los Sistemas de Información Geográfica (SIG) para el análisis espacial de la pobreza en Colombia ha sido poco explorado. Los aportes de este tipo de análisis son:

- La identificación de patrones de distribución espacial, la localización de los conglomerados críticos de pobreza y la identificación de desigualdades socio-espaciales, que permiten focalizar la política pública de acuerdo con las variaciones de los niveles de pobreza en el territorio.
- Conocer el papel de diferentes variables del territorio en la dinámica de la pobreza y el grado en que la interacción entre las características sociales y territoriales influyen en la conformación de conglomerados de bienestar o de pobreza.

En este enfoque del Centro de Estudios Económicos Regionales se tienen dos trabajos. El primero es de Pérez (2005) titulado «Dimensión espacial de la pobreza en Colombia». En este el autor encuentra que existe dependencia espacial de la pobreza medida con el NBI (necesidades básicas insatisfechas) y el ICV (índice de calidad de vida), a nivel municipal y departamental para 1985 y 1993. Concluye que «la pobreza de un municipio depende en forma importante del nivel de pobreza de los municipios vecinos» (Pérez, 2005, 38). Los conglomerados de pobreza encontrados se ubican en la Costa Caribe, en el Chocó y en la zona cafetera.

El segundo es el trabajo de Galvis y Meisel (2010) llamado «Persistencia de las desigualdades regionales en Colombia: Un análisis espacial», el cual confirma los hallazgos de Pérez (2005) en cuanto a la autocorrelación espacial positiva de la pobreza en Colombia y los conglomerados de pobreza en las zonas antes mencionadas. Estos autores hacen un análisis bivariado del índice de Moran, medida de autocorrelación espacial, cruzando los NBI calculados a partir de los censos de población de 1985, 1993 y 2005. Con este análisis los autores buscaron comprender en los tres momentos del tiempo, la dependencia espacial de la pobreza, y la existencia de trampas de pobreza espaciales y efectos de vecindario, las cuales están localizadas en el Caribe, el Pacífico, la Orinoquia y la Amazonia; es decir, que en estas regiones los municipios pobres tendieron a permanecer rodeados de pobres en el tiempo.

Finalmente, Vargas (2012), en su trabajo de grado titulado «Aproximación espacial al estudio de la pobreza en Colombia: un estudio aplicado con información del año 2005», identifica y analiza los patrones espaciales de la pobreza en Colombia usando el ICV. Al igual que en los trabajos anteriores, la autora confirma la autocorrelación espacial positiva y los conglomerados de pobreza en las zonas ya mencionadas. Adicionalmente, la autora realiza una regresión geográficamente ponderada, en la cual busca comprender cómo las variables: distancia a la capital nacional, la tasa de urbanización (razón entre población urbana y población total municipal), la generación de recursos propios (ingresos tributarios/ingresos totales), los ingresos públicos per cápita y la eficiencia en la gestión municipal tienen un efecto sobre la pobreza. Vargas (2012) concluye que la actividad económica se distribuye de manera desigual en el espacio.

3. Metodología

3.1. Datos de pobreza. IPM-2005

El índice de pobreza multidimensional (IPM) fue desarrollado por Alkire y Foster (2007, 2011) y fue adaptado para Colombia por el Departamento Nacional de Planeación (Angulo *et al.*, 2011). Anualmente este es calculado a partir de la Encuesta de Calidad de Vida, con la cual se obtienen resultados representativos a nivel nacional, cabecera y resto. Para calcular el IPM a nivel municipal⁴ se usó el Censo de Población de 2005, y se evaluaron las siguientes cinco dimensiones, de las cuales se desprenden 15 indicadores: condiciones educativas, condiciones de la niñez y juventud, trabajo, salud y servicios públicos y condiciones de la vivienda. Este índice tiene una estructura anidada en donde cada una de las dimensiones pesa 0,2 y cada uno de los 15 indicadores tiene el mismo peso dentro de cada dimensión. Son consideradas como pobres aquellas personas que pertenecen a un hogar que presenta privación en por lo menos el 33% (o 5/15) de los indicadores.

3.2. Análisis espacial

El análisis espacial parte de la Primera Ley de la Geografía, según la cual «todo está relacionado con todo lo demás, pero cosas más cercanas están más relacionadas que cosas distantes» (Tobler, 1970, traducción libre). Este puede ser definido como el conjunto de técnicas estadísticas y matemáticas aplicadas a los datos distribuidos sobre el espacio geográfico (Buzai y Baxendale, 2006), o bien como el estudio cuantitativo de fenómenos que están localizados en el espacio (Bailey y Gattrel, 1995).

El análisis espacial incluye el análisis exploratorio de datos espaciales (AEDE), el cual es un conjunto de técnicas para describir y visualizar

⁴Según el DANE, el resto del municipio corresponde al área que está por fuera del perímetro urbano de la cabecera municipal. Puede estar constituido por centros poblados y población dispersa. Esta entidad geográfica se asocia a la zona rural del municipio.

la distribución espacial de los datos, identificar atípicos espaciales, y conglomerados que indican heterogeneidad o estructuras espaciales en los datos (Chasco, 2001). El AEDE se lleva a cabo por fases, las cuales se describen a continuación.

Primera fase: mapas temáticos.

La primera fase del AEDE incluyó la elaboración de dos tipos de mapas temáticos con el fin de representar la tendencia espacial de los datos. El primero es el mapa de cuantiles, que permite representar la tendencia espacial global del IPM a nivel total, cabecera y resto, agrupando los datos en categorías (cuantiles) con igual número de observaciones (Chasco, 2001). El segundo es el mapa de desviación estándar del IPM, que permite identificar la dispersión de los datos con respecto a la media, en un rango estandarizado, y brinda una primera aproximación de los atípicos espaciales (Chasco, 2001).

Segunda fase: tendencia global de los datos.

En esta fase del AEDE se realiza el análisis de tendencia con la herramienta «Trend Analysis», mediante el cual, el componente de una superficie puede ser modelado mediante una fórmula matemática, con el fin de determinar si existe alguna tendencia direccional de los datos del IPM 2005 y si esta es o no aleatoria.

El gráfico de la herramienta «Trend Analysis»⁵ (gráfico 1) hace posible examinar la tendencia espacial global de los datos en una perspectiva tridimensional. Las ubicaciones de los puntos correspondientes a las entidades geográficas se representan en el plano xy (puntos en color rojo). A cada entidad se le proyecta la altura en el eje z, basado en el

valor que toma la variable de estudio. Luego, cada punto con sus coordenadas (x, y, z) se proyecta en las superficies Occidente-Oriente (plano xz, puntos en color verde oscuro) y Norte-Sur (plano yz, puntos en color azul).

Según la dispersión de los datos, se ajusta una curva de tendencia para los planos Norte-Sur (línea azul) y Occidente-Oriente (línea verde), como se muestra en el gráfico 1. Estas curvas permiten determinar si la tendencia corresponde a un polinomio de primer (línea recta), segundo (trazo en forma de U) o tercer orden (trazo sinuoso). Si la línea fuera plana, esto podría indicar que no hay tendencia. Sin embargo, como se observa en la línea verde del gráfico, se registran valores altos al Occidente del país, que decrecen a medida que nos desplazamos al centro y aumentan al Oriente; esta curva de tendencia es indicativa de un polinomio de segundo orden.

Tercera fase: autocorrelación espacial e identificación de conglomerados de pobreza.

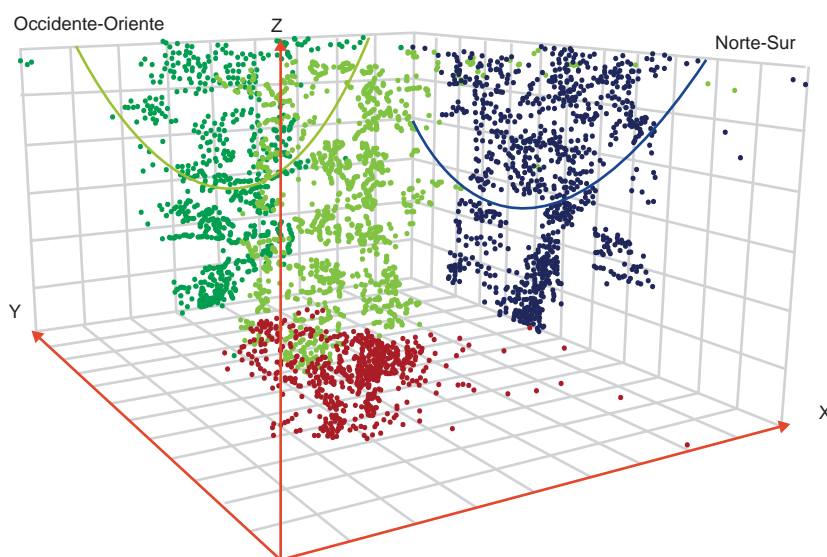
Por último se estudian los efectos espaciales como la autocorrelación

espacial y la heterogeneidad; y se estudian los patrones espaciales con las técnicas de análisis espacial de carácter global y local, con el índice de Moran y el Anselin local de Moran. Los patrones espaciales reconocidos se establecen en la condición estadísticamente significativa (análisis confirmatorio de datos espaciales [ACDE]).

La dependencia o autocorrelación espacial puede definirse como la relación funcional que se da entre el valor de la variable en determinado punto en el espacio y lo que ocurre en lugares vecinos a este. Una variable presenta autocorrelación espacial cuando los valores observados de esta en el espacio dependen de los valores que esta toma en regiones vecinas (Chasco, 2001).

La heterogeneidad espacial, la cual suele presentarse junto con la autocorrelación espacial, puede ser entendida como la manera desigual en la que un fenómeno se distribuye en espacio; ejemplos de esta son las situaciones de centro-periferia, norte-sur, occidente-oriente (Chasco, 2004). Esto se identifica por medio de las diferencias sistemáticas de media,

Gráfico 1. Análisis de tendencia de los datos de IPM 2005, a nivel resto.



Fuente: DANE. Elaboración propia.

⁵ Las herramientas utilizadas en este trabajo hacen parte del software SIG de nombre ArcGIS 10.1 de ESRI, y las conceptualizaciones de relaciones espaciales utilizadas en el análisis espacial de los datos IPM 2005 se describen en Estrada y Moreno (2013), Anexo 2.

⁶ Cómo funciona la autocorrelación espacial (índice de Moran global). Fuente: ArcGIS Resources. URL: <http://help.arcgis.com/es/arcgis-desktop/10.0/help/index.html#//005p0000000t000000> (Consultado el 05/09/2013).

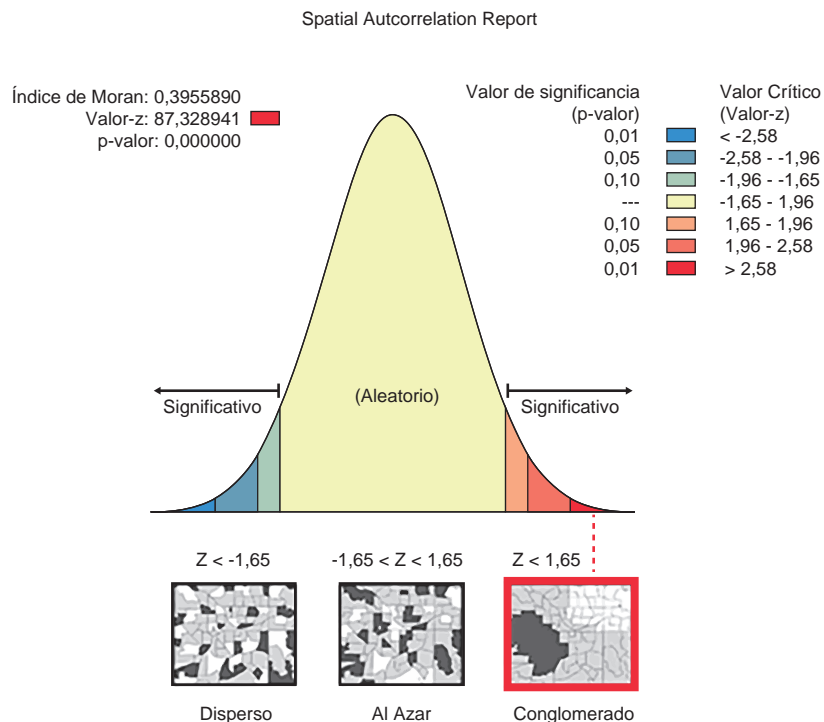
varianza u otros parámetros de un fenómeno en distintas regiones geográficas (Sánchez, 2006).

Se calcula el índice de Moran, que se denota con la letra I, el cual mide la autocorrelación espacial global basada en las ubicaciones y los valores de las entidades simultáneamente. Este índice es una estadística deductiva, que oscila entre -1 y 1 y se interpreta dentro del contexto de la hipótesis nula⁶, la cual establece aleatoriedad espacial completa⁷. Dado que el valor del índice de Moran no se interpreta directamente, se revisa el P valor y el Z crítico. Si de acuerdo con el P valor se rechaza la hipótesis nula, se observa el Z crítico (desviación estándar) para indicar el tipo de patrón espacial que se presenta (dispersión o agrupamiento).

Cuando el Z crítico y el índice de Moran toman un valor positivo quiere decir que existe autocorrelación espacial positiva, que indica que el patrón espacial es de tipo conglomerado; es decir, un valor alto rodeado de valores altos o un valor bajo rodeado de valores bajos (gráfico 2). La negativa corresponde al patrón espacial disperso o uniforme, es un patrón de tipo conglomerado en donde un valor alto o bajo puede estar rodeado indistintamente de valores altos y bajos; y, por último, la autocorrelación espacial inexistente o cero refleja un patrón aleatorio o al azar. Cuando el índice de Moran se acerca a cero esto indica que no hay una relación espacial entre los valores de análisis (Cepal-Unicef, 2012).

En caso de que los datos presenten un patrón espacial de conglomerado, se procede a localizar los conglomerados o concentraciones

Gráfico 2. Prueba de significancia para el índice de Moran.



Fuente: ArcGIS 10.1

espaciales (Hot Spot) de valores similares de la variable de estudio. En otras palabras, se busca localizar los municipios con alto IPM que tienden a estar rodeados de municipios con alto IPM. Esto se puede lograr mediante el cálculo de los estadísticos de G_i^* de Getis-Ord (Getis y Ord, 1992; Ord y Getis, 1995) y el Anselin local de Moran.

El estadístico de G_i^* indica el grado o intensidad de concentración de valores altos o de bajos de una variable en el espacio, identificando los conglomerados estadísticamente significativos. El estadístico Anselin local de Moran detecta no solamente los conglomerados sino también los valores atípicos espaciales⁸; es decir, un valor alto

rodeado por valores bajos o un valor bajo rodeado por valores altos. Cuando se presentan las relaciones estadísticamente no significativas quiere decir que hay unidades territoriales donde el valor de la variable de interés no se relaciona significativamente con los valores que presentan sus vecinos. En el presente estudio se utilizó el estadístico Anselin local de Moran.

La detección de los conglomerados requiere la definición de relaciones espaciales que permita diferenciar cuáles entidades se consideran contiguas o relacionadas entre sí. La conceptualización de las relaciones espaciales entre unidades se puede realizar teniendo en cuenta la distancia entre las unidades o la contigüidad basada en los bordes de los elementos.

Para este estudio, la vecindad entre las unidades territoriales Total y Resto se construyó con una matriz

⁶ Aleatoriedad espacial completa hace referencia a los efectos de primer orden o de igual probabilidad, y de segundo orden o de independencia de los datos.

⁸ Un valor positivo del índice Anselin local de Moran indica que la entidad geográfica está rodeada de entidades cuyos valores son similares, y por tanto, dicha entidad hace parte de un conglomerado. Un valor negativo del índice Anselin local de Moran indica que la entidad geográfica está rodeada de entidades cuyos valores son disímiles, y por tanto, la entidad es un atípico espacial. Fuente: ArcGIS Resources. URL: http://resources.esri.com/help/9.3/arcgisengine/java/Gp_ToolRef/Spatial_Statistics_tools/how_cluster_and_outlier_analysis_colon_anselin_local_moran_s_i_spatial_statistics_works.htm (Consultado el 02/09/2013).

de pesos de contigüidad basados en el criterio «reina»⁹, el cual considera como vecinas las unidades territoriales que se tocan en cualquier parte de los bordes o vértices que las conforman.

Para el caso de las cabeceras municipales fue necesario reevaluar el concepto de contigüidad, dado que las cabeceras limitan en su mayoría con el resto municipal y no con otras cabeceras; por esto, se consideraron como cabeceras vecinas aquellas que estuvieran a una distancia máxima de 314.851

metros. Esta distancia se calculó mediante iteración del cálculo del índice de Moran con el fin de obtener la distancia a la cual la autocorrelación espacial del IPM, a nivel de Cabecera, era la máxima.

4. Resultados

Para el cálculo del IPM a nivel municipal se usó la información del Censo General 2005¹⁰ y se trabajó con información de IPM para 1114 municipios, 1097 cabeceras y 1112 resto municipal (tabla 1).

Tabla 1. Descripción de la información disponible para las entidades geográficas 200511.

Descripción	Total Municipal	Resto	Cabecera
Total entidades 2005	1119	1119	1119
Omisión censal municipal del 100%	5	5	5
Omisión censal del 100% o sin información en Resto		2	
Omisión censal 100% en cabecera v con cartografía	1		1
Cabeceras sin cartografía			6
Corregimientos sin información a nivel cabecera de IPM			10
Total de unidades con información IPM			
Total de unidades de información IPM	1114	1112	1097

Fuente: DANE. Elaboración propia.

Tabla 2. Estadísticos descriptivos del IPM 200512.

	Muestra	Mínimo	Máximo	Rango	Media	Desviación Estándar
IPM Total	1114	14,27	100	85,73	69,47	16,39
IPM Cabecera	1097	13,23	100	36,77	53,54	18,50
IPM Resto	1112	22,99	100	77,01	79,91	13,55

Fuente: DANE. Elaboración propia.

Los estadísticos de resumen o descriptivos (tabla 2) permiten comparar y poner en evidencia los

comportamientos dispares al tratar los datos del IPM para el Total, Cabecera y Resto; el valor máximo

fue 100 para cada uno de los tres niveles. El valor mínimo del IPM Resto es 1,7 veces el valor mínimo del IPM Cabecera y 1,6 veces el del IPM Total.

El promedio del IPM Cabecera es significativamente menor al del IPM Total y Resto, siendo este de 53,50. Sin embargo, estos datos son los que presentan la mayor dispersión. Lo contrario ocurre con el IPM Resto que tiene el promedio más alto pero unos datos más concentrados.

Adicionalmente, se identificó que los municipios que presentan alto porcentaje de omisión censal, presentan también un alto IPM. No obstante, esta relación no se mantiene en el sentido inverso; es decir, existen municipios que con un alto IPM poseen un bajo porcentaje de omisión censal.

4.1. Análisis exploratorio de datos espaciales (AEDE)

A continuación se presentan los resultados para los mapas temáticos, análisis de tendencia, autocorrelación espacial y de conglomerados de pobreza y atípicos espaciales.

4.1.1. Primera fase: mapas temáticos

Mapas de cuantiles. Antes de iniciar el análisis de los mapas de cuantiles, vale la pena resaltar las diferencias en los rangos. Conservando en mente la brecha que existe entre la pobreza a nivel de cabecera y resto, nótese que la media del IPM Total es 69,47 con una desviación estándar de 16,36; el IPM Resto tiene una media de 79,97 con una desviación estándar

⁹El criterio «reina» hace alusión al movimiento que tiene la ficha reina en el ajedrez. Además del criterio, en el análisis espacial se utiliza el orden de contigüidad, donde el primer orden se refiere a los inmediatamente vecinos y segundo, si existe un municipio intermedio.

¹⁰Es importante tener en cuenta que el Censo General 2005 tuvo una no cobertura de 3,7%; de la cual el 1,4% se debió principalmente a dificultades de acceso a los hogares por razones de violencia y el 2,3% restante corresponde a la no cobertura final de unidades censales.

¹¹Los municipios que tuvieron una omisión censal del 100%, tanto en cabecera como en resto, fueron: La Victoria (Amazonas), Mapiripána (Guainía), Pana Pana (Guainía), Morichal (Guainía) y Paoa (Vaupés); los municipios que tuvieron omisión censal 100% o sin información de IPM a nivel de resto fueron: Carurú (Vaupés), que presentó una omisión censal del 100%, y Puerto Rondón (Arauca), respectivamente; el municipio que tuvo omisión censal de 100% a nivel de cabecera fue Mapiripán (Meta); las cabeceras sin cartografía fueron: San Miguel (Putumayo), Mirití-Paraná (Amazonas), Puerto Colombia (Guainía), La Guadalupe (Guainía), Papunaua (Vaupés) y Yavaraté (Vaupés); y los corregimientos departamentales sin información de cabecera fueron: El Encanto (Amazonas), La Chorrera (Amazonas), La Pedrera (Amazonas), Puerto Alegría (Amazonas), Puerto Arica (Amazonas), Puerto Santander (Amazonas), Tarapacá (Amazonas), Barranco Minas (Guainía), San Felipe (Guainía) y Cacahual (Cacahual). En 2005 existían 20 corregimientos departamentales, de los cuales 10 no cuentan con información a nivel cabecera y para el cálculo del IPM se tomaron como resto.

¹²Ver Estrada y Moreno (2013), anexo 1 para mayor detalle de los resultados obtenidos en el análisis exploratorio de datos (AED).

de 13,55 y el IPM Cabecera tiene una media de 53,54 con una desviación estándar de 18,50. Esto se convierte en una dificultad para comparar los tres mapas al no tener los mismos rangos¹³.

En el mapa 1, se observa que a nivel Total los municipios que exhiben mayores niveles de pobreza multidimensional están al sur del país, en la costas Pacífica y Caribe, y en la Alta Guajira. La región andina exhibe los menores niveles de pobreza multidimensional; sin embargo, pueden encontrarse municipios con altos niveles de IPM rodeados de municipios con bajos niveles de IPM. Igualmente resaltan las capitales departamentales como Puerto Carreño (Vichada) y Leticia (Amazonas), cuyo IPM Total se acerca más a los niveles de la región andina, que a los de los municipios más cercanos.

Es importante señalar que la Costa Pacífica presenta un comportamiento heterogéneo, y se evidencia una discontinuidad de los altos niveles de pobreza en el municipio de Buenaventura (Valle del Cauca).

En el mapa 2 se observa que a nivel de Restos, si bien la tendencia es similar a la de Total, se puede constatar que los niveles de pobreza multidimensional son significativamente superiores. Así, en la región andina hay ahora más municipios afectados por los altos niveles de IPM; en la Costa Caribe y La Guajira aumentan los niveles de pobreza.

El resto de las capitales del Vichada y Amazonas muestra un comportamiento similar al de todos los municipios del departamento. Al observar el departamento de Amazonas ve-

mos que los municipios presentan unos niveles de pobreza menores que los municipios de otros departamentos como Putumayo y Caquetá.

En el mapa 3, se muestra el IPM Cabecera y vemos que se conservan las tendencias evidenciadas en los dos mapas anteriores; sin embargo, en este caso resaltan las capitales Barranquilla (Atlántico), Cartagena (Bolívar), Santa Marta (Magdalena), Leticia (Amazonas), Arauca (Arauca), Mitú (Vaupés) y Puerto Inírida (Guainía) con niveles de pobreza multidimensional menores a los de las otras cabeceras cercanas. Así mismo, la cabecera de Buenaventura (Valle del Cauca) presenta un IPM cercano al de las demás cabeceras del departamento del Valle del Cauca.

Este mapa resalta el carácter preeminentemente rural de las regiones Amazonía que agrupa los departamentos de Amazonas, Vaupés, Guainía, Caquetá, Putumayo y Guaviare; y Orinoquia que agrupa Casanare, Arauca, Meta y Vichada. Exceptuando la franja de transición hacia la Cordillera Oriental.

Los tres mapas anteriores demuestran que efectivamente la distribución de la pobreza multidimensional presenta un patrón espacial, coincidente con los trabajos anteriores (Pérez, 2005; Galvis y Meisel, 2010; y Vargas, 2012).

4.1.2. Mapas de desviación estándar

Los mapas de la desviación estándar del IPM para los tres niveles de estudio permiten identificar qué tan alejados de la media, por encima o por debajo, se encuentran el IPM Total, Resto y Cabecera de cada uno de los municipios. La importancia de estos reside en que atenúan los resultados obtenidos anteriormente, en los cuales se identificaba la región andina como aquella con los niveles más bajos de IPM; además resaltan la presencia de conglomerados de

bienestar alrededor de ciertas capitales y municipios del país.

En el mapa 4, se observa que el IPM Total de las capitales Bogotá y Bucaramanga, y de los municipios Envigado, La Estrella, Sabaneta, Rionegro e Itagüí en Antioquia, y Duitama en Boyacá, se encuentra a más de 2,5 desviaciones estándar de la media nacional, la cual fue de 69,5 (tabla 2). El IPM Total de Manizales, Armenia, Pereira, Medellín y Cali está entre 2,5 y 1,5 desviaciones estándar por debajo de la media nacional. Alrededor de estos hay municipios que presentan un IPM Total por debajo o cercano a la media nacional.

Esto muestra la existencia de polos de bienestar en el interior de la región andina, entendidos como aquellos municipios que presentan un IPM Total a más de 1,5 desviaciones estándar por debajo de la media nacional y que en su primer anillo de contigüidad los municipios tienen un IPM Total cercano a la media nacional o incluso inferior.

Llaman la atención Barranquilla, Cartagena, Neiva, Ibagué y Popayán con desviaciones estándar del IPM Total entre 2,5 y 1,5 por debajo de la media, pero los municipios aledaños tienen unas desviaciones estándar por encima de la media nacional.

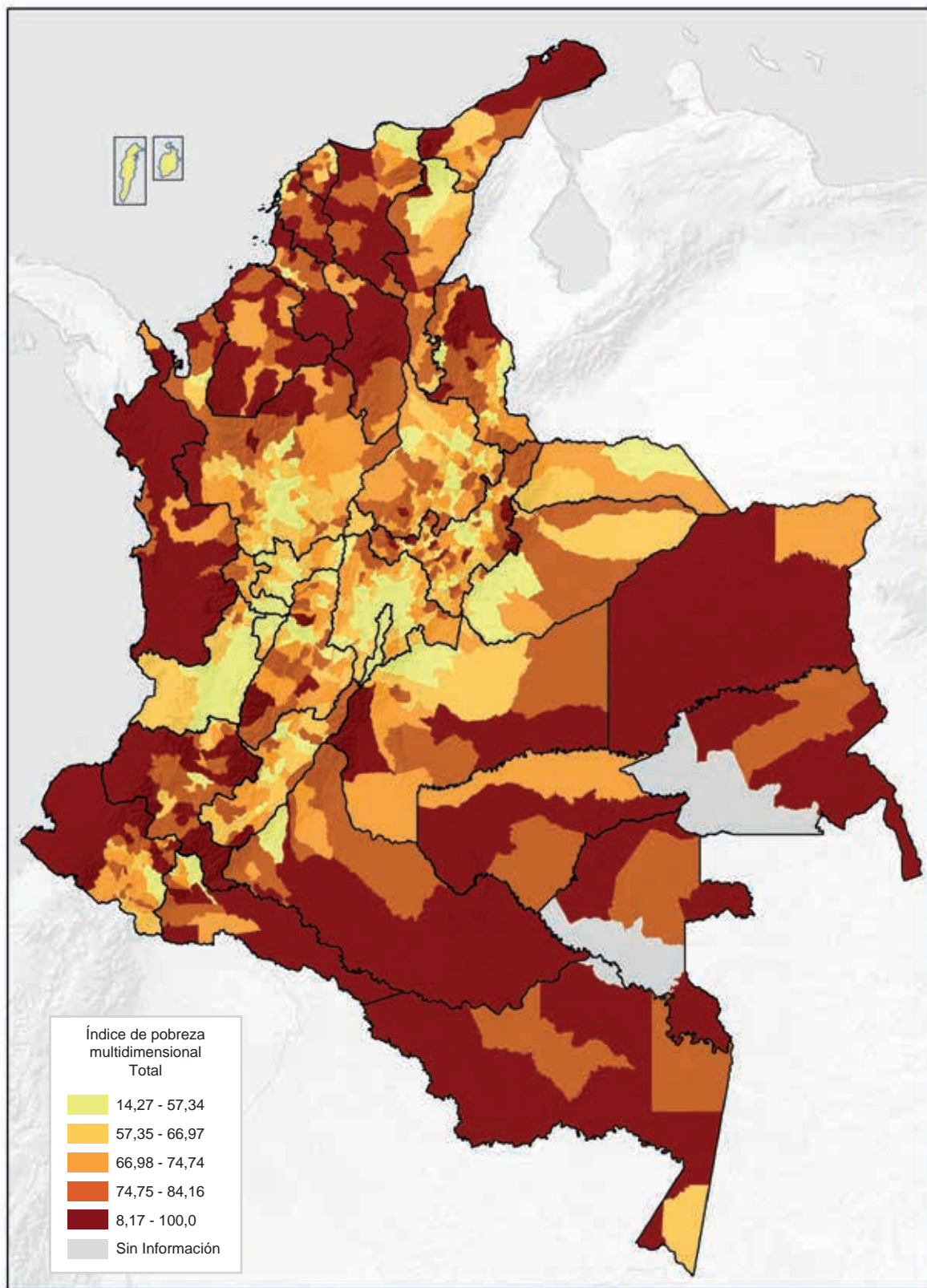
En el mapa 5 se puede apreciar que en las áreas Resto existen también polos de bienestar, pues hay grupos de municipios que presentan un IPM Resto a más 1,5 desviaciones estándar por debajo de la media nacional, la cual fue de 79,9 (Tabla 2). Contrario a lo visto en el IPM Total, en Resto solo las capitales Cali y Medellín están a 2,5 desviaciones estándar por debajo de la media. En este caso, son los municipios aledaños a las capitales los que presentan un IPM Resto dentro de este rango, y se encuentran principalmente en Antioquia¹⁴, Cundinamarca¹⁵

¹³ Dado que la distribución de los datos de pobreza difiere considerablemente a nivel Total, Cabecera y Resto, no es posible utilizar los mismos rangos de clasificación en los mapas temáticos elaborados; adicionalmente no existe una clasificación temática de la pobreza que pueda ser utilizada para clasificar los municipios estudiados.

¹⁴ En el caso de Antioquia, los municipios son: La Estrella, Sabaneta, Envigado, Rionegro y El Retiro.

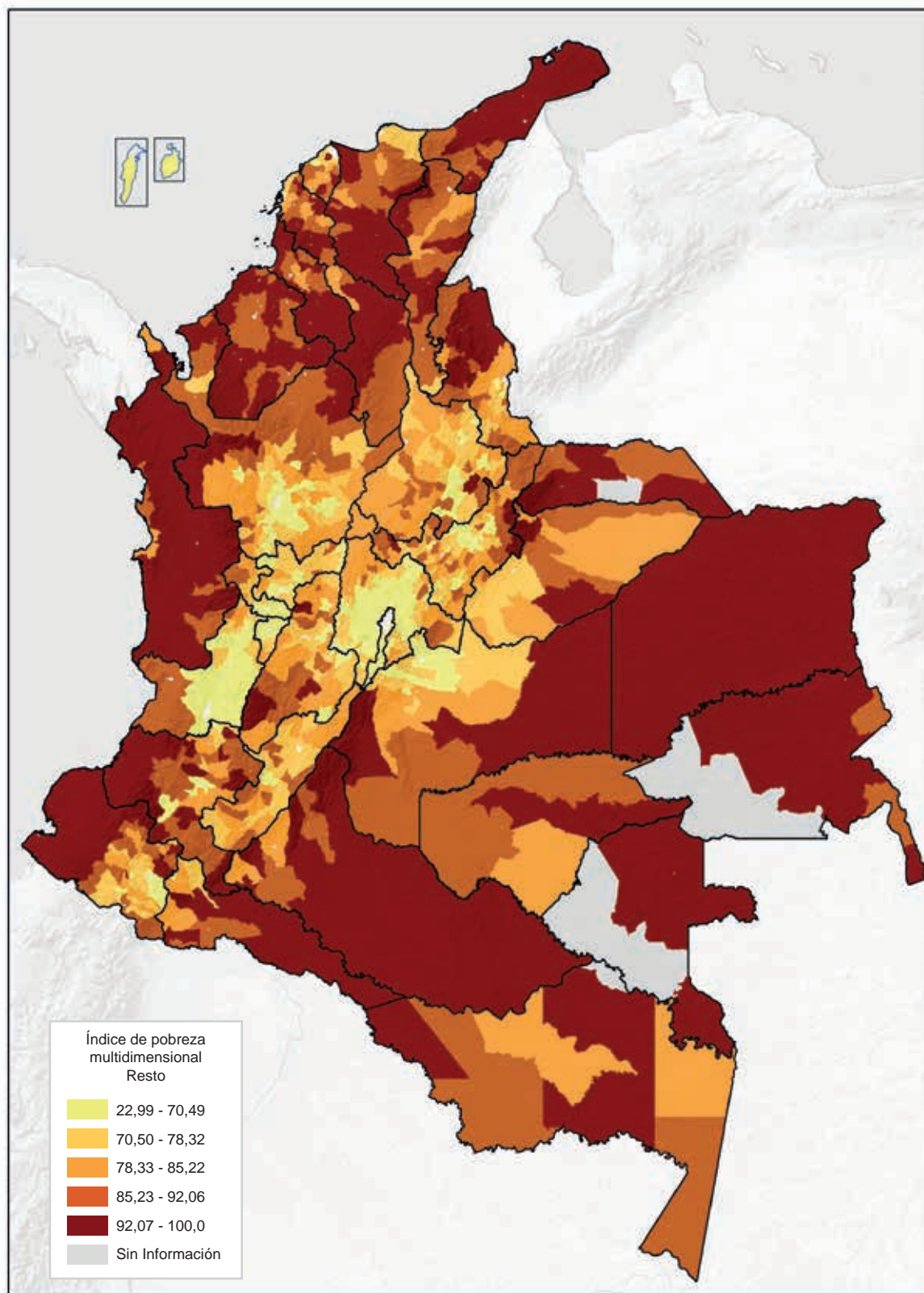
¹⁵ En el caso de Cundinamarca son: Granada, Mosquera, Madrid, Tenjo, Cota, Subachoque, Tabio, Chía, Cajicá, Sopó, Guasacá, Tocancipá, Gachancipá, Sesquillé y Cogua.

Mapa 1. Mapa de cuantiles del IPM 2005, a nivel municipal.



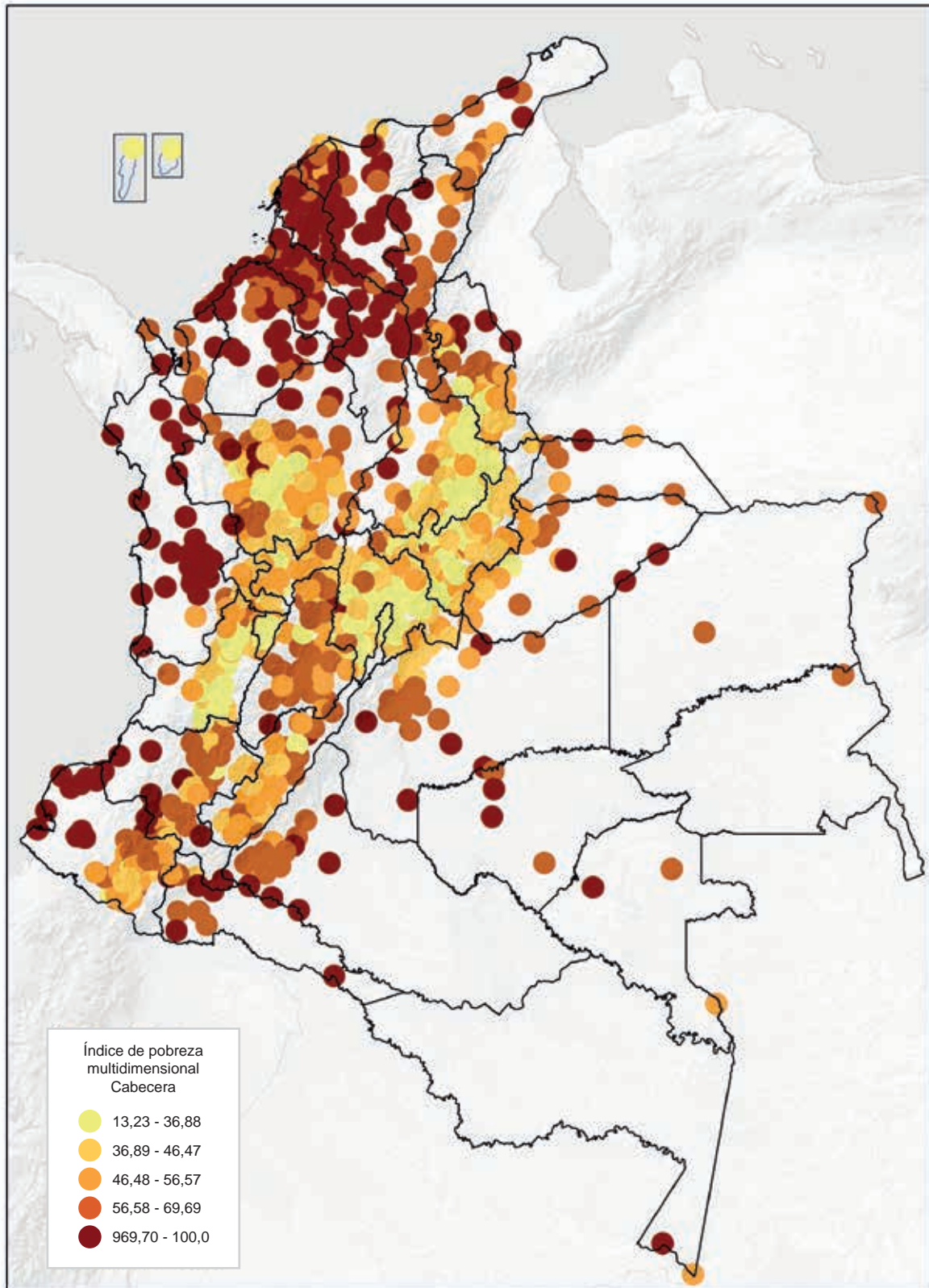
Fuente: DANE. Elaboración propia.

Mapa 2. Mapa de cuantiles del IPM 2005, a nivel resto.



Fuente: DANE. Elaboración propia.

Mapa 3. Mapa de cuantiles del IPM 2005, a nivel cabecera.



Fuente: DANE. Elaboración propia.

y Valle del Cauca¹⁶. Ahora, entre 2,5 y 1,5 desviaciones estándar por debajo de la media nacional se encuentran las capitales Manizales, Armenia y Pereira.

Llama la atención Antioquia que presenta dos comportamientos bastante disímiles entre el norte y el sur del departamento. Al norte se observa que la pobreza en el área rural tiene un comportamiento similar a la del Chocó y los departamentos de la Costa Caribe; y al sur del departamento la pobreza es similar a la de la región andina. Algo similar sucede en el departamento del Valle del Cauca, cuyos niveles de pobreza rural en la parte occidental son similares a los de los municipios del Chocó y del occidente de Nariño, y en la parte oriental tiene valores como los del Eje Cafetero; mostrando continuidad de la pobreza en la Costa Pacífica.

Llama la atención el comportamiento casi homogéneo de la pobreza rural en los departamentos del Huila, Santander, Cundinamarca, Caldas, Risaralda y Quindío, que se encuentran cercanos a la media nacional (69,5%).

En el mapa 6, se puede apreciar que el IPM Cabecera de la región andina está significativamente por debajo de la media nacional, la cual fue de 53,5 (tabla 2). En este caso se presenta más un fenómeno de corredor de bienestar que de polo de bienestar; entendiendo el corredor de bienestar como el conjunto de municipios que presentan un IPM Cabecera a más de 1,5 desviaciones estándar por debajo de la media nacional y se agrupan de manera elongada.

Vemos que hay dos corredores: uno que se extiende desde el suroriente del departamento del Valle del Cauca hasta el sur del departa-

mento de Antioquia; y otro desde el sur del departamento de Cundinamarca pasando por Boyacá hasta el suroriente de Santander.

Al igual que en los mapas anteriores, queda en evidencia que si bien las capitales Barranquilla (Atlántico), Cartagena (Bolívar), Santa Marta (Magdalena), Leticia (Amazonas), Mitú (Vaupés), Puerto Carreño (Vichada) y los municipios del sur del departamento de La Guajira presentan un IPM menor con respecto a la media, no hacen parte de los corredores de bienestar.

Con los mapas de desviaciones estándar se pudo detectar la existencia de polos de bienestar a nivel de IPM Total e IPM Resto; y de corredores de bienestar a nivel de IPM Cabecera; indicando dos tipos de patrones espaciales que caracterizan la pobreza multidimensional en Colombia.

4.1.3. Segunda fase: análisis de curvas de tendencia norte-sur y occidente-oriente.

Los gráficos de análisis de tendencia muestran la tendencia de una variable en las direcciones norte-sur y occidente-oriente. En el gráfico 3 se observa la tendencia del IPM para el total municipal. Los puntos en color rojo representan cada uno de los municipios del universo de estudio; así, Colombia corresponde al plano xy que es la base del cubo.

El valor registrado de IPM para cada municipio corresponde al valor en el eje z, y cada uno de estos valores es proyectado en dos planos, como se explicó en la metodología. En el plano xz están proyectados los datos en el sentido Occidente-Oriente, así la nube de puntos en color verde oscuro y su línea de tendencia muestra que el IPM Total toma valores altos al Occidente; a medida que atravesamos el centro del país vemos que disminuye la pobreza multidimensional, y al acercarnos al Oriente

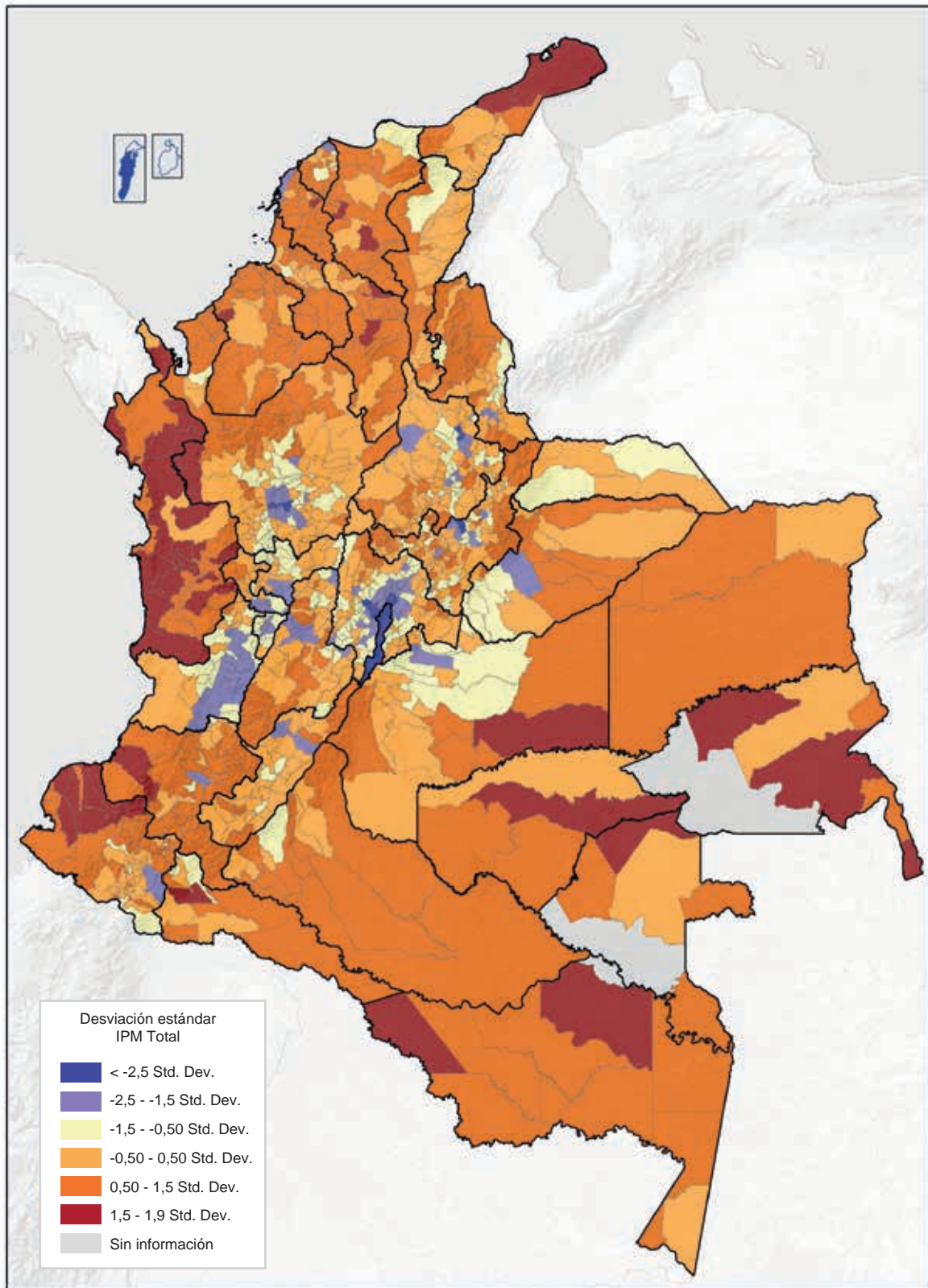
esta vuelve a aumentar. Aunque se confirman los resultados obtenidos anteriormente en torno a la región andina, lo que se puede apreciar con esta tendencia es que los municipios del occidente presentan niveles de pobreza más altos que los del oriente colombiano.

En el plano yz están proyectados los datos en el sentido norte-sur, representados con una nube de puntos en color azul y una línea de tendencia, y muestran que el IPM Total toma valores altos al Norte; a medida que atravesamos el centro del país este disminuye, y aumentan nuevamente hacia el sur del país. Al igual que en la tendencia occidente-oriente se confirman los resultados anteriores frente al centro del país, y en el sentido norte-sur podemos decir que los municipios que se encuentran al norte del país son tan pobres como aquellos que están al sur, en donde hay municipios cuyo IPM Total tomó un valor de 100. Esto hace que la concavidad de la curva de tendencia sea más pronunciada que la registrada en la dirección occidente-oriente.

En el gráfico 4 está representada la tendencia del IPM Resto. La línea de tendencia en el plano xy muestra que el IPM Resto toma los valores más altos posibles tanto al Occidente como al Oriente del país, revelando una marcada diferencia con el centro; se confirma la brecha significativa de la pobreza multidimensional registrada en la zona rural del centro del país, y la registrada en la zona oriental y occidental respectivamente. La línea de tendencia en el plano yz indica que si bien el IPM Resto toma valores altos al Norte, las zonas rurales de los municipios del sur del país están en peores condiciones con relación a las del Norte. En este caso, la tendencia norte-sur nos está indicando que el IPM Resto exhibe un comportamiento menos desigual entre el norte y el centro

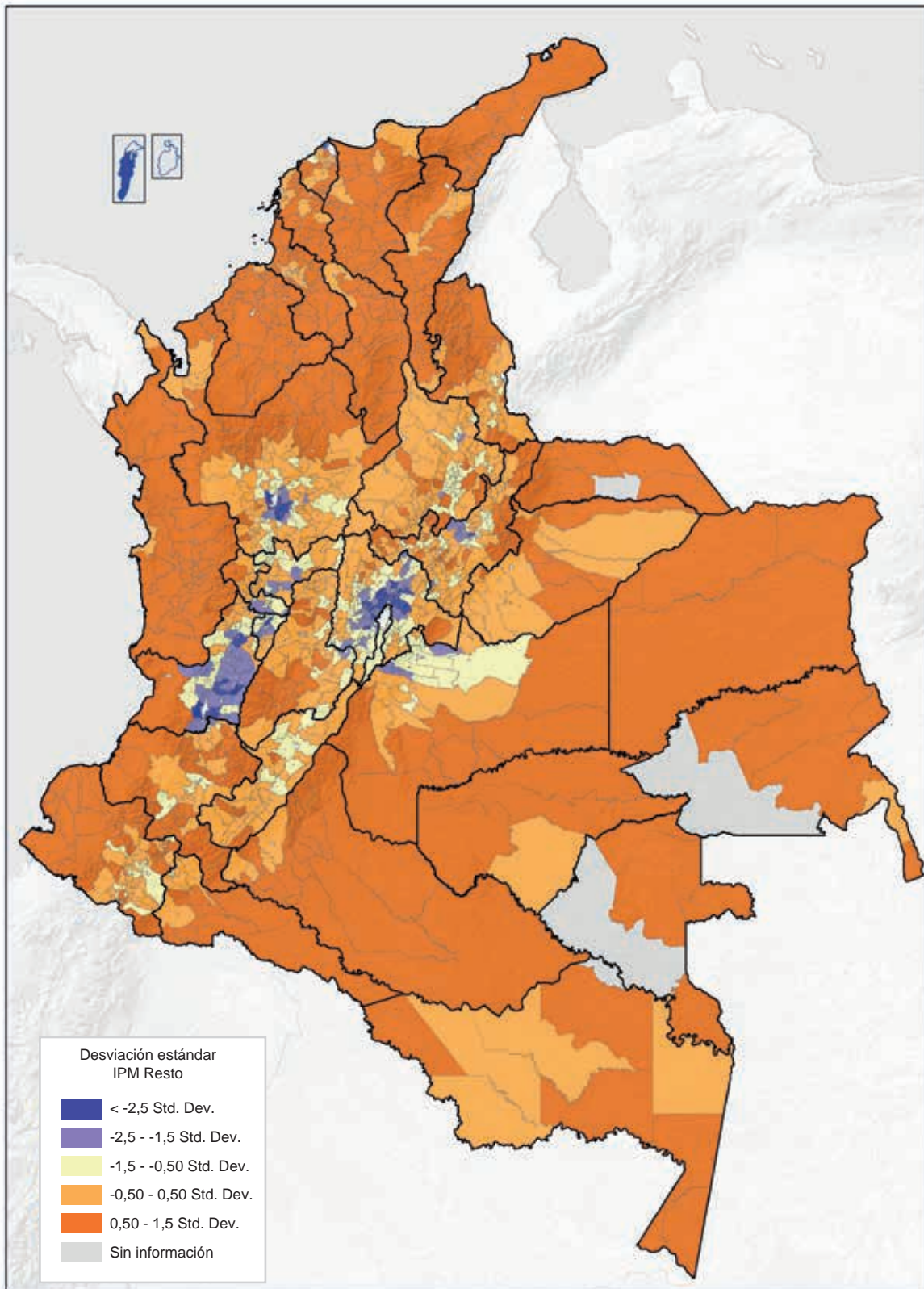
¹⁶ En el caso del Valle del Cauca son: Guacarí, Cerrito, Candelaria y Zarzal.

Mapa 4. Mapa de desviación estándar del IPM 2005, a nivel municipal.



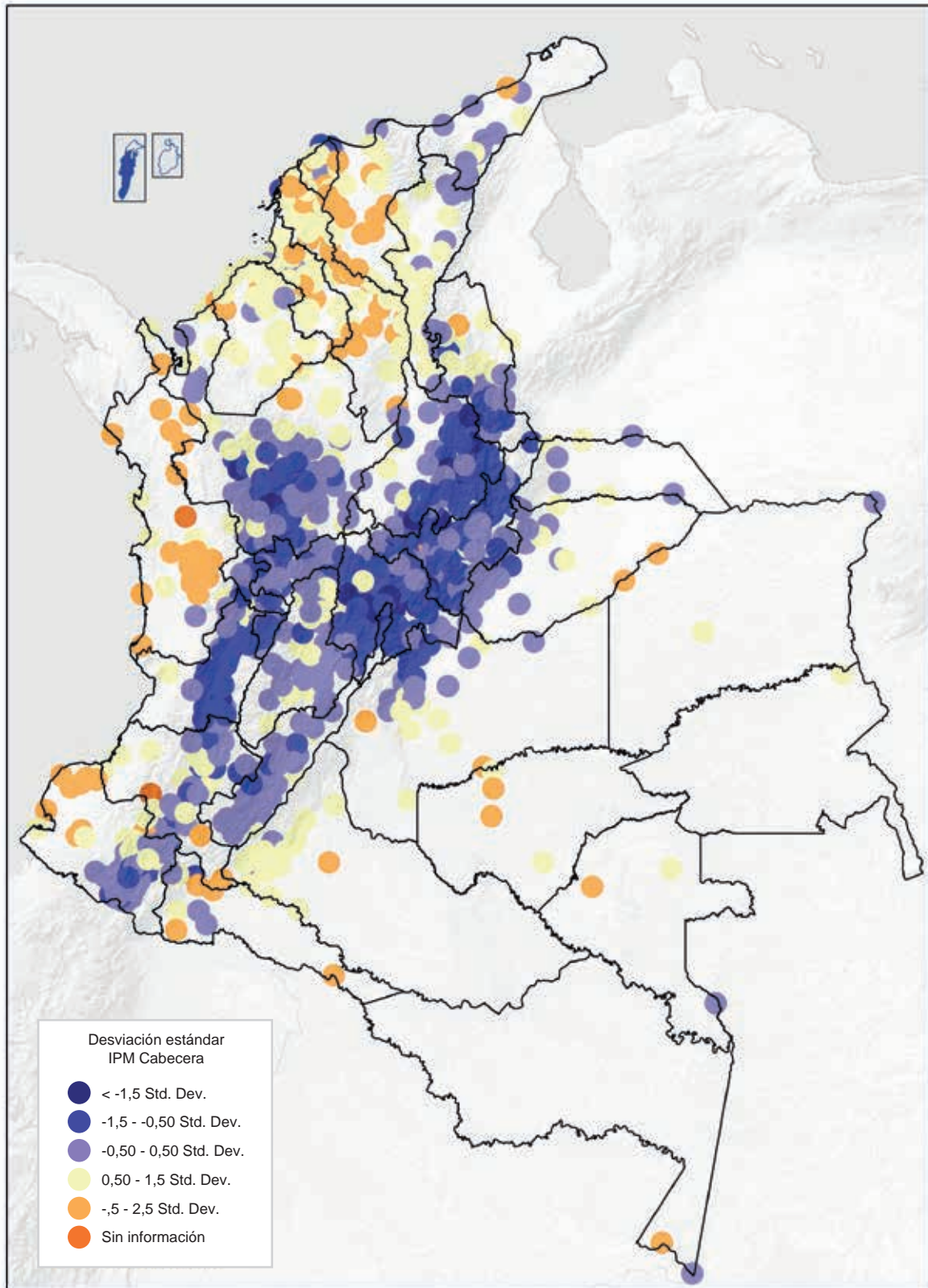
Fuente: DANE. Elaboración propia.

Mapa 5. Mapa de desviación estándar del IPM 2005, a nivel de resto.



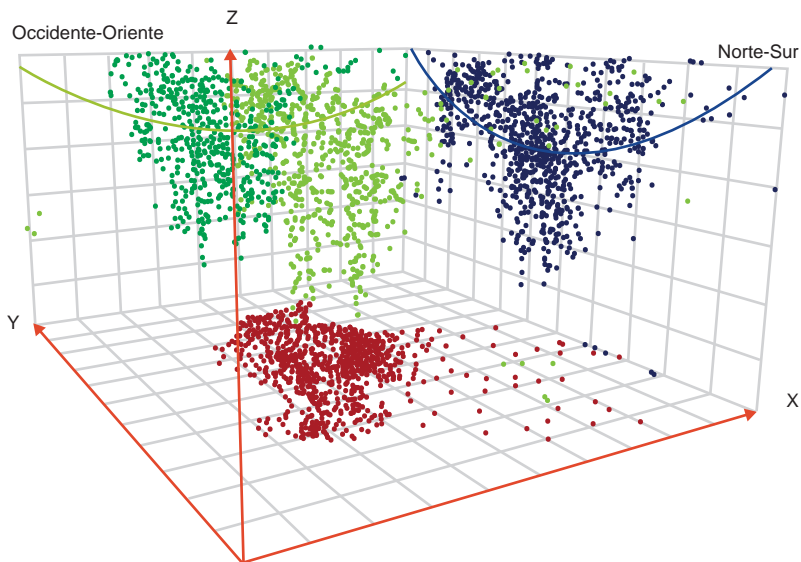
Fuente: DANE. Elaboración propia.

Mapa 6. Mapa de desviación estándar del IPM 2005, a nivel de cabecera.



Fuente: DANE. Elaboración propia.

Gráfico 3. Análisis de tendencia de los datos de IPM 2005, a nivel municipal.



Fuente: DANE. Elaboración propia.

del país; pero muy desigual entre el centro y el sur del país.

En el gráfico 5 está representada la tendencia del IPM Cabecera. La línea de tendencia en el plano xz muestra que las cabeceras municipales del occidente del país presentan mayores niveles de pobreza multidimensional que aquellas del oriente del país. Así la situación de las cabeceras del oriente del país con respecto a las del centro es menos desigual, que aquellas del occidente y estas últimas. La línea de tendencia en el plano yz muestra que el IPM Cabecera toma los valores máximos tanto al norte como al sur del país, presentándose una alta diferenciación con la dinámica del centro. Es importante recordar que las cabeceras al sur del país no son numerosas.

Globalmente, el análisis de tendencia del IPM muestra que a nivel Total y de Cabeceras, el oriente del país se encuentra en mejores condiciones que el occidente; mientras que el norte del país presenta una menor pobreza multidimensional a nivel de IPM resto con respecto al sur. En los demás casos se observó que el oriente es tan pobre como el occidente en sus zonas rurales; y el sur es tan pobre como el norte para el total municipal y las cabeceras. El centro siempre presenta los valores más bajos de IPM Total, Cabecera y Resto. Esto se modela mediante una curva polinomial de segundo orden en los tres niveles, que se puede representar como una «U», tanto en la dirección norte-sur como occidente-oriente.

4.1.4. Tercera fase: autocorrelación espacial y conglomerados de pobreza¹⁷

Índice de Moran

Basados en los resultados anteriores, se espera encontrar una autocorrelación positiva y un patrón

espacial de tipo conglomerado de la pobreza multidimensional en Colombia. Para comprobar estadísticamente esta hipótesis se calcula el índice de Moran, que como se mencionó en la metodología, permite encontrar si hay autocorrelación espacial en torno a la pobreza multidimensional; es decir: si los municipios con altos niveles de pobreza tienden a estar rodeados de municipios con altos niveles de pobreza, y viceversa (agrupamiento); si la pobreza multidimensional está dispersa a nivel municipal, o tiene un patrón aleatorio.

Se confirma que la pobreza multidimensional presenta autocorrelación espacial positiva y estadísticamente significativa para los niveles a nivel municipal total, cabecera y resto (tabla 3); por cuanto, a partir del P-Valor, se rechaza la hipótesis nula de aleatoriedad en la distribución de la pobreza multidimensional en los tres niveles de estudio a un nivel de confianza de 99%. También se observa que tanto el índice de Moran como el Z crítico tomaron valores positivos, lo que indica que en los tres casos la pobreza multidimensional presenta un patrón de conglomerado (cluster)¹⁸ en los valores altos y en los valores bajos de la pobreza multidimensional.

Conglomerados de pobreza

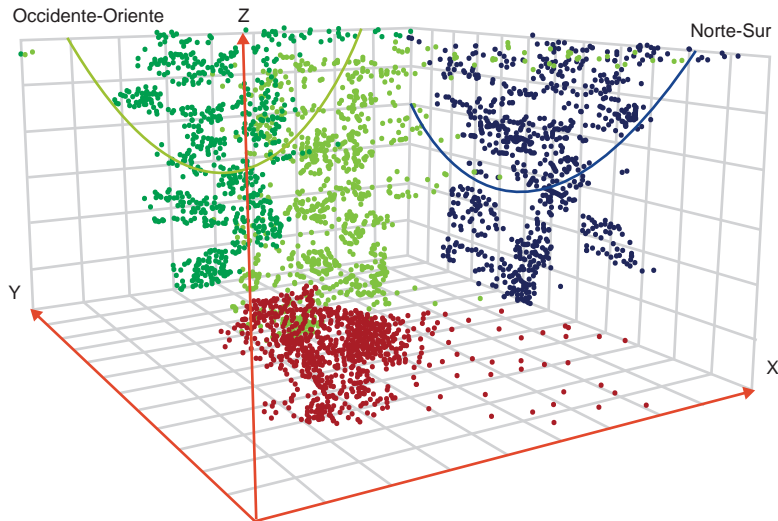
Habiendo comprobado la existencia de autocorrelación espacial positiva de la pobreza multidimensional en Colombia, se calcula el estadístico de Anselin local de Moran para localizar las concentraciones estadísticamente significativas de alto o bajo IPM y los atípicos espaciales —entidades con IPM bajos rodeados de IPM altos, y viceversa—.

En el mapa 7 y el mapa 8, se puede observar que para el IPM Total y Resto el fenómeno de concentración que más se presenta en el país es el de altos niveles de pobreza.

¹⁷ El departamento Archipiélago de San Andrés, Providencia y Santa Catalina no se pudo tener en cuenta en esta fase porque no existe continuidad espacial con la porción continental de Colombia.

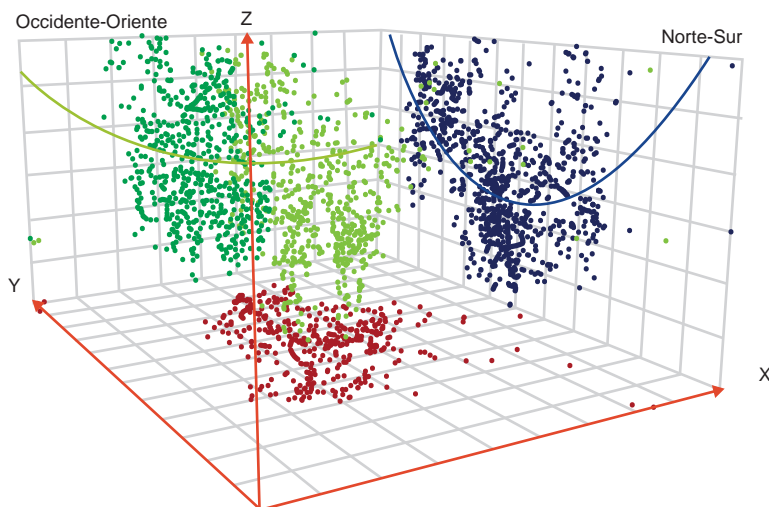
¹⁸ En este caso no se puede comparar el valor que tomó el índice de Moran a nivel Cabecera con los resultados para Total y Resto por la diferencia en la conceptualización de las relaciones espaciales (véase Metodología).

Gráfico 4. Análisis de tendencia IPM 2005, a nivel resto.



Fuente: DANE. Elaboración propia.

Gráfico 5. Análisis de tendencia IPM 2005, a nivel cabecera.



Fuente: DANE. Elaboración propia.

Se registran algunas concentraciones de bajo IPM Total y Resto en el región andina, y algunos casos de atípicos espaciales¹⁹.

En cuanto a los conglomerados es interesante anotar cómo estos su-

peran las fronteras departamentales. Podemos identificar en el corredor del Pacífico dos conglomerados de municipios con alto IPM Total y Resto, que son interrumpidos al nivel del Valle del Cauca y el sur del Chocó. Estos municipios que no hacen parte del conglomerado corresponden a zonas no significativas estadísticamente para el estadístico de Anselin local de Moran; lo que significa que hay presencia de

municipios donde el valor del IPM no se relaciona significativamente con los valores que registran los municipios aledaños.

En el caso de IPM Total, algunos municipios que son frontera entre Antioquia y Chocó forman conglomerados de alto IPM; e igualmente sucede al norte en los departamentos de Córdoba, Sucre, Bolívar, Magdalena y Cesar. Sin embargo, al observar el IPM Resto la concentración de alta pobreza multidimensional se amplía formando un claro corredor que une los departamentos de Chocó, Antioquia y Córdoba.

Al sur del país se presenta este comportamiento entre los departamentos de Putumayo, Caquetá, Amazonas y Guaviare para ambos niveles. El conglomerado de alta pobreza a nivel resto entre Putumayo, Caquetá y Amazonas es menos extenso con relación al nivel total, ya que los municipios La Chorrera, Puerto Santander, Puerto Arica y El Encanto en el departamento del Amazonas se registran como zonas estadísticamente no significativas. Adicionalmente hay un cambio significativo en el departamento del Vichada donde la concentración continúa hacia el Guaviare y el Vaupés.

Resalta el departamento del Atlántico que hace parte marginalmente de las concentraciones de alta pobreza de sus departamentos contiguos y, por el contrario, exhibe un pequeño conglomerado de bienestar en el municipio de Puerto Colombia, que debido a la escala es difícil de observar. En lo que respecta el departamento de La Guajira vemos que la concentración de pobreza multidimensional aumenta en el Resto con respecto al IPM Total.

Los conglomerados de bajo IPM Total y Resto confirman la existencia de polos de bienestar relacionados con las capitales departamentales –Cali, Medellín, Bogotá, Manizales, Armenia, Pereira, Bucaramanga y

¹⁸ En este caso no se puede comparar el valor que tomó el índice de Moran a nivel Cabecera con los resultados para Total y Resto por la diferencia en la conceptualización de las relaciones espaciales (véase Metodología).

¹⁹ Véase Estrada y Moreno (2013), Anexo 3, para una lista detallada de los conglomerados y atípicos espaciales.

Tabla 3. Índice de Moran para IPM 2005: Total, Resto y Cabecera.

I de Moran Global			
	IPM Total	IPM Resto	IPM Cabecera
Índice de Moran	0,602393	0,692594	0,39589
Índice Esperado	-0,000898	-0,0009	-0,000912
Varianza	0,000335	0,000338	0,000021
Z crítico	32,98314	37,748742	87,328941
P-Valor*	0,000000	0,000000	0,000000

Fuente: DANE. Elaboración propia.

Villavicencio— y sus municipios alrededores. En el caso del oriente del departamento del Valle del Cauca su dinámica de bienestar se une a la de los departamentos de Caldas, Risaralda y Quindío, formando un corredor de bienestar que no había sido observado en los mapas de desviación estándar.

Los atípicos que resultan a nivel Total confirman estadísticamente lo que se había observado anteriormente en el caso de Cartagena, mostrando que presenta un IPM bajo rodeado de IPM altos. Contrario a esto y que no se había apreciado tan claramente, se encuentra el atípico al sur del departamento del Tolima y frontera con el Valle del Cauca, en el municipio de Rio Blanco, el cual registra un alto IPM rodeado de bajos IPM.

En el mapa 9, se puede observar que para el IPM Cabecera el fenómeno de concentración de alta y baja pobreza es más homogéneo, en el sentido en que se presentan ambos fenómenos en amplias zo-

nas del país. Adicionalmente, los dos tipos de casos de atípicos espaciales aumentan con relación a lo encontrado para IPM Total y Resto.

Ahora bien, observamos que se mantiene el fenómeno de concentración de la alta pobreza en Nariño y Cauca; sin embargo, esta vez se extiende hacia el Caquetá y Putumayo. Así como también se confirma la existencia de los corredores de bienestar anteriormente descritos.

Se encuentran ocho atípicos espaciales de bajo IPM rodeados de alto IPM que corresponden a: Barranquilla, Puerto Colombia y Soledad en el Atlántico, Cartagena, Valledupar, Santa Marta, La Playa en Norte de Santander, y San Pedro de Cartago en Nariño. Los atípicos espaciales de alto IPM rodeados de bajo IPM, se localizan principalmente en los departamentos de Antioquia, Santander, Meta, Tolima, Cundinamarca, Boyacá, Casanare y Arauca²⁰. Estos atípicos espaciales podrían ser sujetos de estudio más detallado con el fin

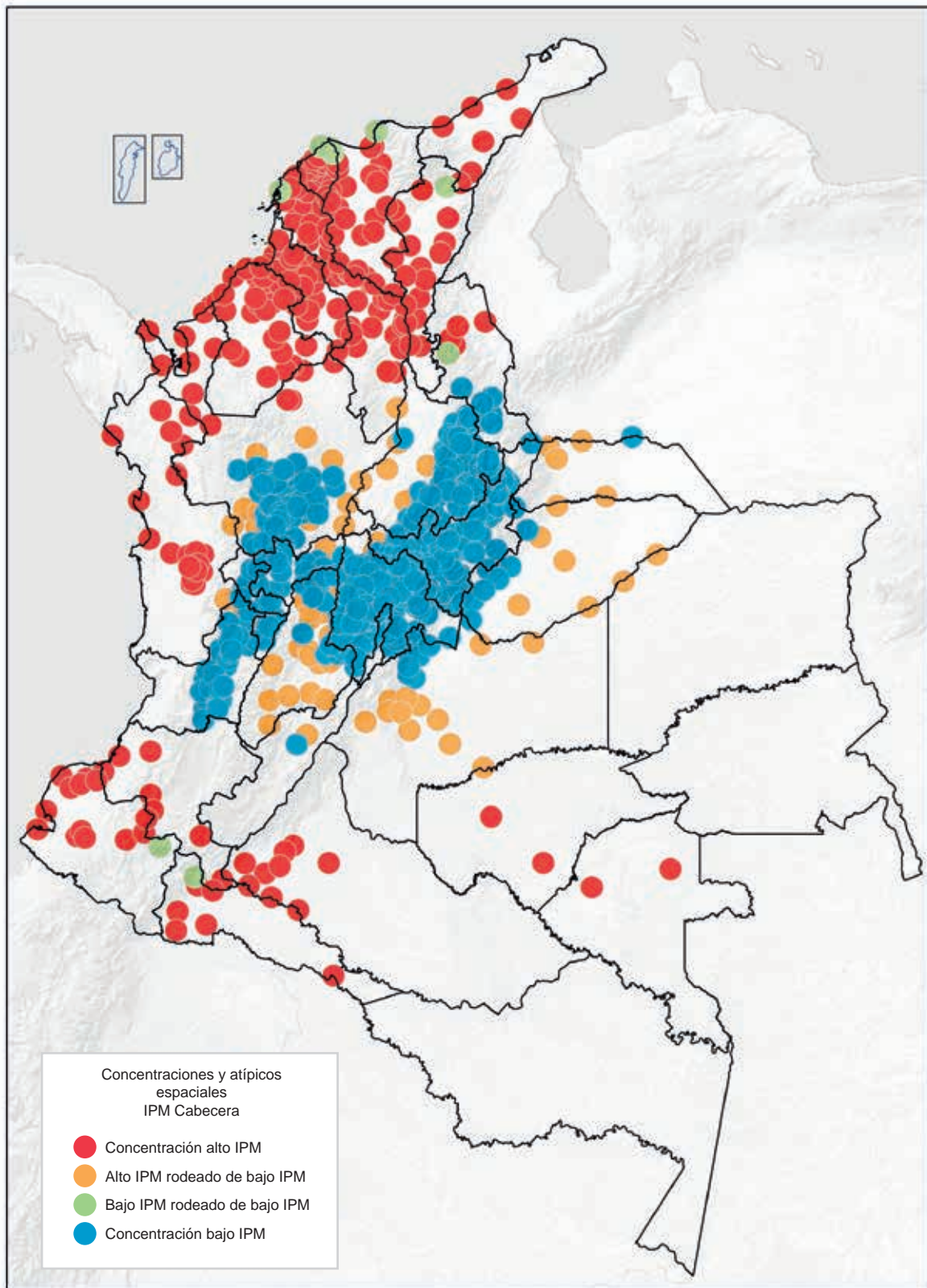
de identificar si temporalmente se consolidan como conglomerados de bienestar o de pobreza.

En cuanto a la dinámica Chocó-Antioquia-Córdoba, se observa que el corredor de altos niveles de pobreza, se mantiene pero esta vez se ubica al norte del departamento de Antioquia, al nivel del Urabá antioqueño. El conglomerado de alta pobreza se refuerza en las cabeceras de los municipios de Córdoba, Sucre, Atlántico, Bolívar, Magdalena, Cesar, Riohacha y al norte del Norte de Santander.

El estadístico Anselin local de Moran permitió identificar conglomerados de alto o bajo IPM, así como atípicos espaciales; para los tres niveles de estudio se observa que estos o se intensifican o se debilitan. El conglomerado de pobreza entre los departamentos de Cauca y Nariño para IPM Total y Resto se amplía hacia los departamentos del Caquetá y Putumayo para IPM Cabecera. El conglomerado de pobreza de la Costa Caribe de IPM Cabecera es más extenso que para IPM Total y Resto. Se confirman las tendencias de los datos y la existencia de polos y corredores de bienestar, encontradas anteriormente. Se encuentran atípicos espaciales que no se habían apreciado en mapas anteriores, en los departamentos del Tolima, Norte de Santander, Nariño, Putumayo y Cesar.

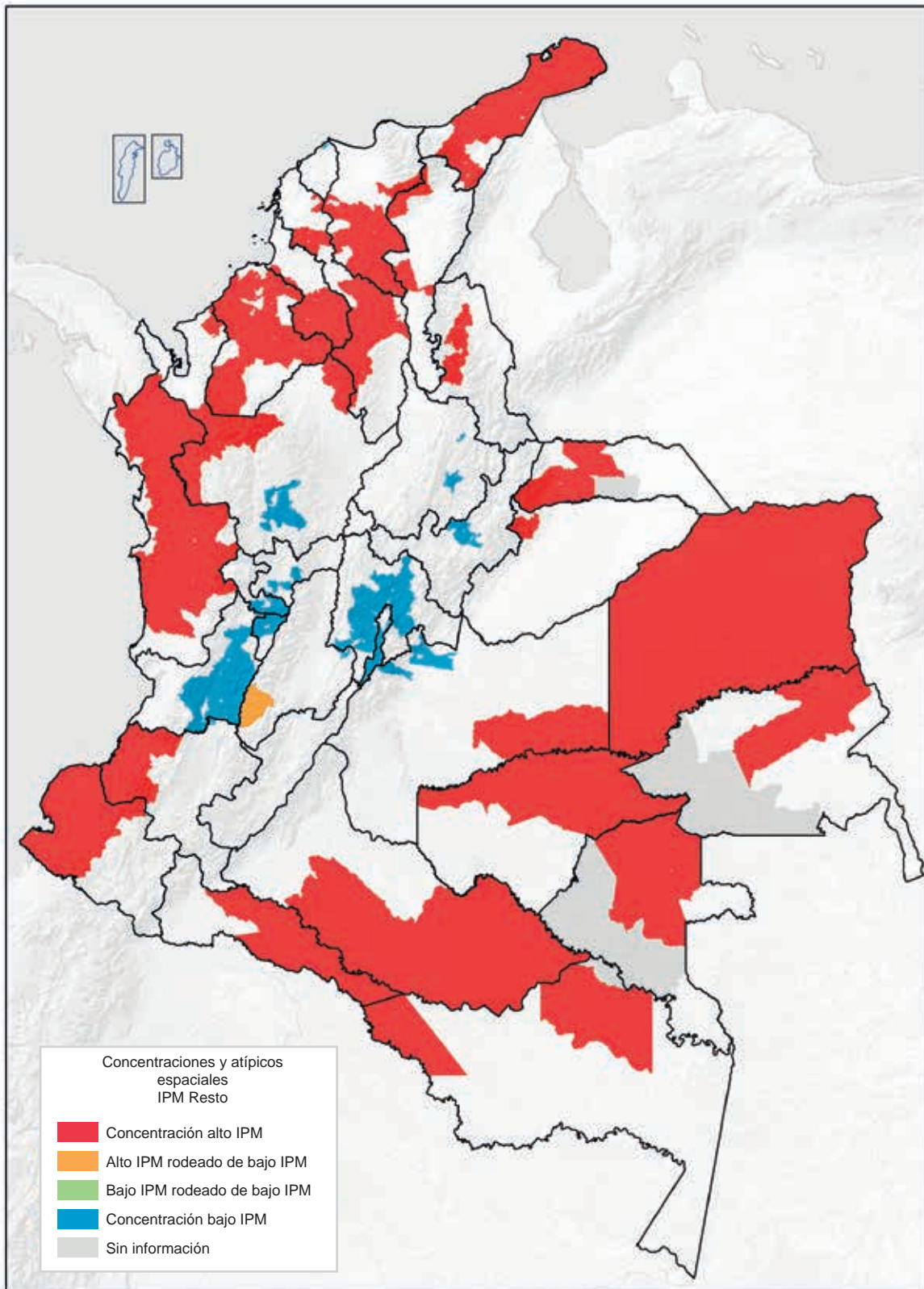
²⁰Véase Estrada y Moreno (2013), Anexo 3.

Mapa 7. Concentraciones y atípicos espaciales para el IPM 2005, a nivel municipal.



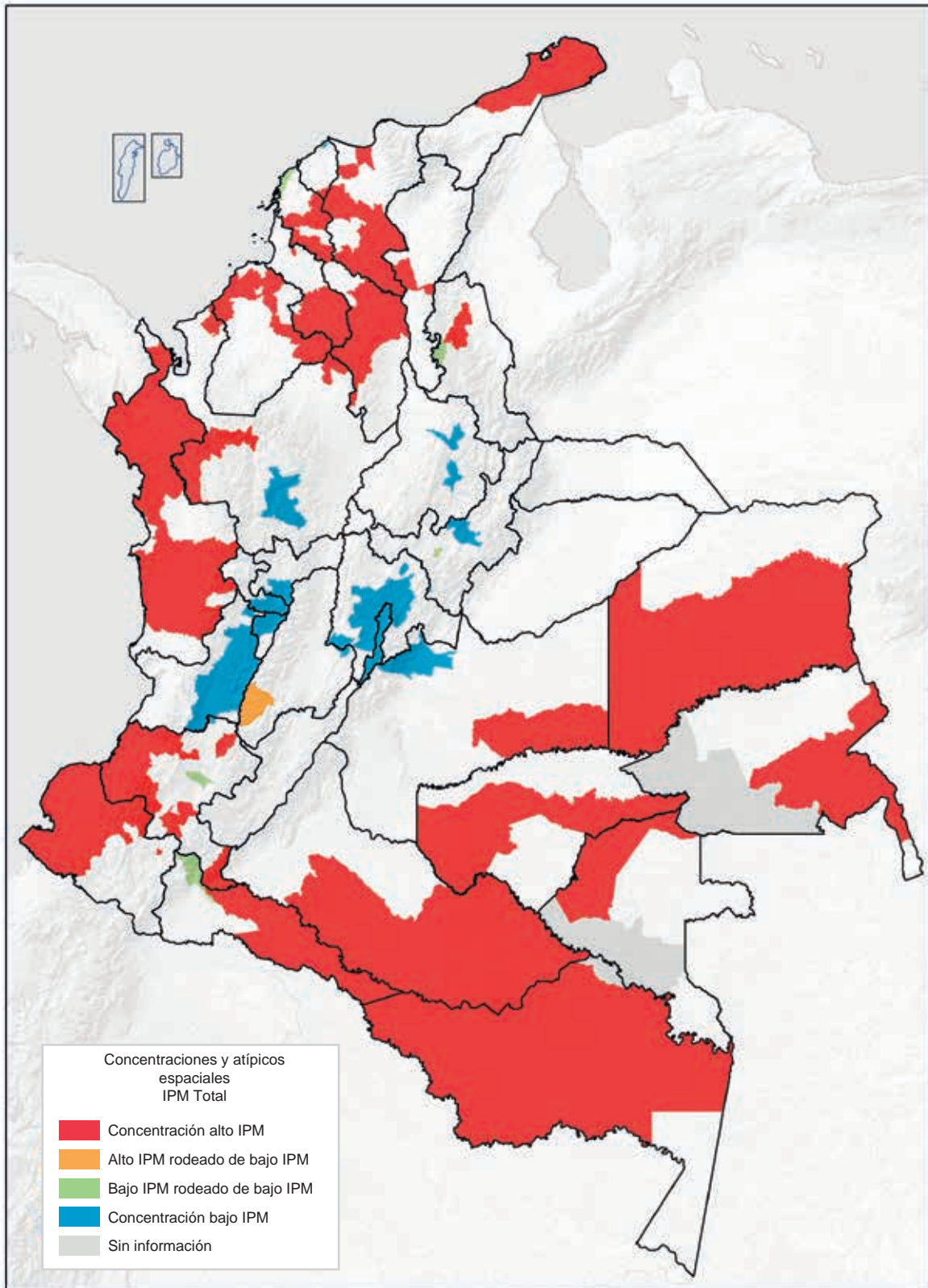
Fuente: DANE. Elaboración propia.

Mapa 8. Concentraciones y atípicos espaciales para el IPM 2005, a nivel de resto.



Fuente: DANE. Elaboración propia.

Mapa 9. Concentraciones y atípicos espaciales para el IPM 2005, a nivel de cabecera.



Fuente: DANE. Elaboración propia.

Conclusiones

A través de los diferentes mapas, gráficos y pruebas estadísticas se comprobó que existe la dimensión espacial de la pobreza multidimensional en Colombia. En otras palabras, la pobreza no se distribuye de manera aleatoria en el espacio sino que presenta un marcado patrón espacial. Se confirman los hallazgos de estudios anteriores (Pérez, 2005; Galvis y Meisel, 2010; Vargas, 2012) en cuanto a que la pobreza de un municipio no solamente depende de su propia condición, sino también de lo que sucede en los municipios aledaños. Sin embargo, se encontraron importantes diferencias en los patrones espaciales según el nivel que se esté estudiando (Total, Cabecera y Resto).

Para todos los niveles se evidenció que la distribución espacial de la pobreza en Colombia supera los límites departamentales, lo cual implica que la concentración de la pobreza no está dada por características y relaciones administrativas sino adicionalmente por condiciones regionales y del territorio; este punto es central para las implicaciones de política territorial, la cual debe interactuar con la región y tener en cuenta las características del territorio que superan las fronteras administrativas.

El análisis espacial con base en el IPM 2005 permitió la identificación de polos de bienestar para los niveles Total y Resto y corredores de bienestar para el nivel Cabecera, asociados a las principales capitales: Bogotá, Cali, Medellín, Bucaramanga, entre otras; o de municipios cercanos a estas. Esto llevaría a pensar sobre la importancia de multiplicar y fortalecer los centros de consumos como estrategia para elevar los niveles de bienestar de la

población. En cuanto a los corredores de bienestar, estos se extienden desde el suroriente del departamento del Valle del Cauca hasta el sur del departamento de Antioquia; y otro desde el sur del departamento de Cundinamarca pasando por Boyacá hasta el suroriente de Santander. Este análisis también sirvió para detectar los casos de Barranquilla, Santa Marta, Cartagena, Popayán, Leticia, Puerto Carreño y Arauca, que si bien presentan bajo IPM, no son polos de bienestar para los municipios aledaños.

Por otro lado, el uso del índice Anselin local de Moran confirmó de manera estadísticamente significativa las apreciaciones sobre los mapas temáticos. También, este estadístico permitió identificar casos atípicos espaciales en los departamentos del Tolima, Norte de Santander, Nariño, Putumayo y Cesar. Más allá de la identificación de los conglomerados de pobreza, resulta de gran importancia recalcar que la continuidad en el espacio de la pobreza supera las fronteras departamentales; además se resalta la heterogeneidad que existe en el interior de los mismos departamentos. Así, recordemos los casos de continuidad de la pobreza multidimensional al occidente de los departamentos del Cauca y Nariño; Putumayo, Caquetá y Amazonas; y Chocó-Antioquia-Córdoba.

Es igualmente interesante el análisis de tendencia y ajuste polinomial de segundo orden, que no había sido explorado por otros autores, pues ofreció una visión comparada de la pobreza en las direcciones norte-sur y occidente-oriente. Gracias a esto se pudo identificar que los niveles de pobreza a nivel de Total y Cabecera tienden a

ser superiores al occidente que al oriente del país; mientras que en la dirección norte-sur la pobreza multidimensional a nivel de Resto es mayor al sur que al norte. Esto se modela mediante una curva polinomial de segundo orden en los tres niveles, que se puede representar como una «U», tanto en la dirección norte-sur como este-oeste.

A manera de apertura a lo que será una segunda etapa de este trabajo, la modelación espacial de la pobreza por cada uno de sus componentes, así como acompañada por otras variables, puede ayudarnos a comprender mejor las dinámicas territoriales de la pobreza multidimensional, identificando qué variables económicas, sociales, demográficas o ambientales inciden en la concentración de pobreza o de bienestar, y encontrar respuestas a preguntas tales como: ¿Por qué algunas ciudades, pese a ser capitales, no participan de los polos y corredores de bienestar? ¿Qué características promueven la consolidación de los denominados corredores de bienestar? ¿Qué explica la dinámica transfronteriza de la pobreza? ¿Cuáles son sus implicaciones en términos de política pública?

Así mismo, la brecha entre la pobreza a nivel resto y cabecera lleva a cuestionarse sobre el impacto diferenciado espacialmente que pueden tener las variables socioeconómicas como la educación, la salud, el mercado laboral, la infraestructura, la descentralización fiscal; y las ambientales como la biodiversidad, el clima, la topografía, la calidad del suelo, etc. Por ejemplo, cabe preguntarse cuál es la influencia que tienen en la pobreza la composición étnica de la poblaciones, las carreteras principales entre los centros

de consumo o en el resto municipal, las actividades extractivas como el petróleo o la minería, y las relaciones transfronterizas con Venezuela, Ecuador o Brasil, entre otras.

La comprensión de estos aspectos fortalecería el diseño, la ejecución y la evaluación de la política pública con enfoque territorial, relanza el debate sobre el concepto de región,

ampliamente utilizado en el lenguaje común y académico, y genera una apertura hacia una medición de la pobreza multidimensional adaptada a las dinámicas territoriales.

Bibliografía

Anker, Richard; Chernyshev, Igor; Mehran, Farhad y Ritther, Joshep (2003). La medición del trabajo decente con indicadores estadísticos. *Revista Internacional del Trabajo*, 122(2).

Bonnet, F., Figuereido, J., Standing, G. (2003). "A family of decent work indexes." *International Labor Review* 142(2).

Camacho, Claudia; Dussán, Luis Felipe y Guataquí, Juan Carlos (2012). Calidad del trabajo en Bogotá: Una aproximación desde el enfoque de Trabajo decente. *Cuadernos de desarrollo económico*, núm. 16. Bogotá D. C.

De justicia. Centro de Investigaciones Socio-jurídicas. (2008). Propuesta sobre la definición de la política pública para la generación de trabajo digno y decente en el país. Universidad de los Andes.

Departamento Administrativo Nacional de Estadística, DANE (2013). *Boletín de Informalidad y Seguridad Social*. Colombia.

De la Garza, E. (2010). Hacia un concepto ampliado de trabajo. Barcelona: Anthropos.

Espinoza, María (2004). *Trabajo decente y protección social*. OIT. Santiago de Chile.

European Commission (2001). *Employment and Social Policies: A*

framework for investing in quality. Communication from the comisión to the council the European Parliament, the Economic and Social Committee. (com 2001). Final Luxemburgo.

Farné, Stefano (2002). Estudio sobre la calidad del empleo en Colombia. *Cuadernos de trabajo* número 3. Observatorio del mercado de trabajo y seguridad social. Universidad Externado de Colombia.

Ferreira, María Teresa (2012). Informal versus Precarious Work in Colombia: Conceptualisation and Operationalisation to Approach Quality of Labour. *International Institute of Social Studies*.

Gamero, Julio (2010). *Empleo y trabajo decente a un año del TLC con EE. UU.* Centro Peruano de Estudios Sociales (CEPES). Perú.

Gittleman, Maury and Howell, David (abril 1995). Changes in the structure and quality of jobs in the United States: effects by race and gender, 1973-1990. *Industrial and Labor Relations Review*, (48)3.

Gruenberg, Barry (1980). The happy worker: Analysis of educational and occupational differences in determinants of job satisfaction. *American Journal of Sociology*. Universidad de Chicago.

Infante, Ricardo y Vega-Centeno, Máximo (diciembre, 2001). La calidad del empleo: lecciones y tareas.

Revista del departamento de Economía. Pontificia Universidad Católica del Perú. XXIV(48), 179-236.

Judt, Tony (2010). Algo va mal. Editorial Taurus.

Levaggi, Virgilio (2006). Democracia y trabajo decente en América Latina. Oficina Internacional del Trabajo. Lima, Perú.

Moore, Henrietta (diciembre, 1995). The future of work. *British journal of industrial relations*, 657-678.

Olmedo, Clara (2006). Flexibilización e institucionalización de la precariedad-informalidad laboral: La experiencia de la provincia de Rioja. *Revista Austral de Ciencias Sociales*, (10).

Organisation for Economic Co-operation and Development, OECD (2009). http://www.canadainternational.gc.ca/g20/summit-commet/g20/declaration_092509.aspx

Organización Internacional del Trabajo (1999). Trabajo decente. Memoria del Director General a la 87.ª reunión de la Conferencia Internacional del Trabajo. Ginebra.

Organización Internacional del Trabajo (2008). Medición del trabajo decente. Documento de debate para la Reunión tripartita de expertos sobre la medición del trabajo decente. Ginebra, 8-10 de septiembre de 2008.

- Organización Internacional del Trabajo*, OIT e *International Labour Organization*, ILO (2011). Políticas y regulaciones para luchar contra el empleo precario. Instituto Internacional de Estudios Laborales. Génova.
- Organización Internacional del Trabajo*, OIT (2012). Informe sobre trabajo en el mundo 2012. Mejores empleos para una economía mejor. Instituto Internacional de Estudios Laborales. Génova.
- (2013). Página oficial sobre trabajo decente. <http://www.ilo.org/global/topics/decent-work/lang-es/index.htm>
- (2013_1). Tendencias mundiales del empleo 2012. Prevenir una crisis mayor del empleo. Instituto Internacional de Estudios Laborales. Génova.
- Pacheco, Edith; De La Garza, Enrique y Reygadas, Luis (2010). Trabajos atípicos y precarización del empleo. Presentación. El Colegio de México.
- Peñas, Ramón (2009). More and better jobs: Conceptual framework and monitoring indicators of quality of work and employment in the EU policy arena. Working paper REC_WP 06/2009.
- Polivka, Eteward Anne (octubre, 1996). Contingent and alternative work. Monthly review arrangement, defined. *Monthly labor review*.
- PGN (2011). Trabajo digno y decente en Colombia. Seguimiento y control preventivo a las políticas públicas. Colombia.
- Reygadas, Luis (2009). Trabajos atípicos, trabajos precarios: ¿Dos caras de la misma moneda? El Colegio de México.
- Rodgers, Gerry y Reinecke, Gerhard (1998). La calidad del empleo: perspectivas y ejemplos de Chile y Brasil. En OIT, Reestructuración, integración y mercado laboral. Crecimiento y calidad del empleo en economías abiertas.
- Rodgers, Gerry y Rodgers Janine, eds. (1989). Precarious Jobs in Labour Market Regulation. The growth of Atypical Employment in Western Europe. Génova.
- Rodgers, Gerry (2007). El trabajo decente como una meta para la economía global. *Revista Futuros*, V(17).
- _____(Julio, 2007_1). Labour market flexibility and decent work. *Working paper* núm. 47. International Labour Organization.
- Rosenthal, Nail (1989). More than wages at issue in job quality debate. *Monthly Labor Review*, 112(12), 4-8.

Submission of articles

Revista ib de la Información Básica Estadística¹

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

Revista ib is an annual publication, refereed by peer reviewers, focused on dissemination of all information related to research activities in order to enhance data sharing processes as well as methodologies associated with the statistical, economic and demographic activities.

Articles must correspond to one of the following activities:

- A. Research projects in statistics, economics or demographics with their unpublished results.
- B. Process analysis, interpretation or criticism concerning statistical, economic and demographic topics based on original sources.
- C. Compendium of analysis, systematization and integration of a minimum number of researches in statistics, economics and demographics that have already been published.

1 ■ General procedure:

1. After receiving the article, the author will receive an email notification and a confirmation that the article has met the overall reporting requirements.
2. In the next stage, the article will be reviewed either by the experts from the National Administrative Department of Statistics - DANE or external experts in order to evaluate the article. This process guided by the editorial committee.
3. The arbitration process will be completely anonymous.
4. During the arbitration process, there will be a permanent communication with the authors to make changes to the text.

¹ Revista ib is a journal that provides the public with basic statistical information, published by the CANDANE (DANE's Statistical Training Center).

Submission of articles

Revista ib de la Información Básica Estadística

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

1. General procedure:

5. In the event of a disagreement among reviewers of the article, a third reviewer will be required.
6. Each article included in the journal must be approved by at least two thematic experts.

2. Basic requirements for the submission of articles to the Revista ib:

Length

Maximum length 20 pages, including tables, figures and references.

Format

Arial 11 point, single spacing and margins of 2.5 inches. Tables and graphs should be submitted in a separate file in Excel.

Requirements

Submission of an article will be held to imply that the material in the article has not been published, is not being published or considered for publication elsewhere.

Summary

The article must contain a maximum of 250 words. It must be submitted in both English and the original language of the text. Both versions should contain a list of at least 3 keywords.

Content

The text should contain the following parts: introduction, literature review or theoretical framework, conclusions and bibliography. In the case of texts that correspond to statistical research, projects should also include materials, methods, results and discussion.

2 ■ Basic requirements for the submission of articles to the Revista ib:

Citation and bibliography

The citation format should be APA. For more information, please consult <http://flash1r.apa.org/apastyle/basics/index.htm>

3 ■ Submission of articles:

- The main text should be well integrated, divided into sections, with descriptive headlines. Whenever possible, avoid footnotes and unnecessary separation of parts of the text such as annexes or appendices. In case where footnotes are necessary, ensure that their role is clarify a point and that they are numbered consecutively.
- The language should be clear, concise and easy to understand for a general reader.
- When using technical terms, make sure that you include a simple definition in parentheses, commas or any other sign that will highlight the term and ensure its clear interpretation. When abbreviations or symbols are included, they must be explained the first time they are named.
- Management of numbers: Numbers should be separated by a point when referring to miles, and with a comma if a number contains decimals. This should be applied within the text, tables, charts and graphs.
- The tables, charts, graphs, equations, etc., should be referenced in the text of the article and placed as close as possible to where they were first cited.
- All tables, charts, graphs, equations, etc., should be numbered consecutively.
- Charts and Tables: Charts shows final values, horizontally or vertically. If it does not sum up or show final values, then it is a table. The title should be in Arial 10 and the body (the part that contains the data) should be in size of 9 points. Nobox, shading, or vertical lines are to be submitted. They should be numbered in a consecutive order and Arabic numerals.
- Graphics: All graphics should be numbered consecutively in Arabic numerals, independent of the numbering of tables, charts and diagrams. No box or shading will be accepted; all axes should have a name and units of measure, without bold, abbreviations or symbols. Number labeling should not be included.
- All graphs, charts and tables should cite the source at the foot of the same, in Arial 8 point.

Submission of articles

Revista ib de la Información Básica Estadística

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

4 ■ Notes:

The authors are responsible for the intellectual property, the use of contents, figures, tables, photographs, etc.

When a document is accepted for publication, the copyright thereof becomes a property of DANE.

The content of each article is entirely the responsibility of the author and not binding by the National Administrative Department of Statistics (DANE), or the Revista ib.

To submit items to the journal or for more information contact:
cavendanop@dane.gov.co

**For information on CANDANE, see the rest of our publications
and teaching program on: <http://www.dane.gov.co/candane>**



Instrucciones para el envío de artículos

Revista ib de la Información Básica Estadística

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

La Revista ib es una publicación periódica, arbitrada por pares evaluadores, orientada a difundir trabajos enmarcados en actividades de investigación para crear un espacio de difusión de datos y metodologías asociados a la actividad estadística, económica y demográfica.

Los artículos deben corresponder a alguna de las siguientes actividades:

- A. Proyectos de investigación en estadística, economía o demografía que produzcan resultados inéditos.
- B. Procesos de análisis, interpretación o crítica sobre temas estadísticos, económicos y demográficos basados en fuentes originales.
- C. Compendios sobre análisis, sistematización e integración de un número mínimo de investigaciones en estadística, economía y demografía ya publicadas.

1 ■ Procedimiento general:

1. Al recibir el artículo, se notificará por correo electrónico al autor y se comprobará el cumplimiento de los requisitos generales de presentación.
2. Posteriormente entrará en un proceso de revisión por parte de expertos temáticos del Departamento Administrativo Nacional de Estadística (DANE) o externos que valorarán el artículo en un procedimiento guiado por el comité editorial.
3. El proceso de arbitraje será totalmente anónimo.
4. Durante el proceso de arbitraje se mantendrá permanente comunicación con los autores para realizar modificaciones al texto.

Instrucciones para el envío de artículos

Revista ib de la Información Básica Estadística

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

1 ■ Procedimiento general:

5. En el caso en que haya discrepancia entre los evaluadores del artículo, se recurrirá a un tercer evaluador.
6. Cada artículo incluido en la Revista ib debe ser aprobado por al menos dos expertos temáticos.

2 ■ Requisitos básicos para la presentación de artículos para la Revista ib:

Extensión

Máximo 20 páginas, incluyendo tablas, gráficos y referencias bibliográficas.

Formato

El artículo se debe entregar en formato Word, letra Arial 11 puntos, interlineado sencillo y márgenes de 2,5 centímetros. Las tablas y los gráficos deben enviarse en archivo aparte en el programa Excel.

Requerimientos

Los artículos enviados no deben haber sido publicados anteriormente, ni sometidos a consideración por otras publicaciones.

Resumen

Debe contener un máximo de 250 palabras y presentarse en inglés y español, e incluir en ambas versiones una lista de mínimo 3 palabras clave.

Contenido

El texto debe contener las siguientes partes: introducción, revisión de literatura o marco teórico, conclusiones y bibliografía. En el caso de los textos que corresponden a proyectos de investigación en estadística deben incluir también: materiales y métodos, resultados y discusión.

2 ■ Requisitos básicos para la presentación de artículos para la Revista ib:

Citación y bibliografía

El formato de citación debe ser APA. Puede encontrar mayor información en <http://flash1r.apa.org/apastyle/basics/index.htm>

3 ■ Presentación de los artículos

- El texto principal debe estar bien integrado, aunque dividido en secciones, mediante encabezamientos de carácter informativo. Deben evitarse, en lo posible, las notas al pie de página y la separación innecesaria de partes del texto en forma de anexos o apéndices. En caso de presentarse, las notas de pie de página deben ser de carácter aclaratorio y estar numeradas consecutivamente.
- El lenguaje debe ser claro, preciso y de fácil comprensión para el lector no especializado.
- Los términos técnicos deben ir seguidos de una definición sencilla entre paréntesis, entre comas o cualquier signo que lo destaque y permita su interpretación.
- Cuando se incluyan siglas o símbolos, estos deben explicarse la primera vez que se nombran.
- Manejo de cifras: deben estar separadas por un punto cuando se trata de miles, y por una coma cuando son decimales. Esto se aplica tanto en el texto, como en las tablas, cuadros y gráficos.
- Las tablas, cuadros, gráficos, ecuaciones, etc. deben ser referenciados en el texto del artículo y ubicarse lo más cerca posible del lugar donde se les cite por primera vez.
- Todas las tablas, cuadros, gráficos, ecuaciones, etc. deben ir numerados consecutivamente.
- Cuadros y tablas: el cuadro debe sumar, ya sea horizontal o verticalmente; si no suma en ninguno de los dos sentidos, o en ambos, es una tabla. El título debe presentarse en letra Arial 10 y el cuerpo (la parte que contiene los datos) debe ir en tamaño de 9 puntos. Deben presentarse sin recuadro, sin sombreados, ni líneas verticales. Van numerados(as) en orden consecutivo.
- Gráficos: todos los gráficos deben tener una numeración consecutiva, independiente de la numeración de las tablas, los cuadros y los diagramas. No se debe presentar el gráfico con recuadro ni sombreados; los ejes deben llevar nombre y unidades de medida, sin negrilla y sin abreviaturas o símbolos; no se deben incluir los rótulos de las cifras.
- En todos los gráficos, cuadros y tablas se debe citar la fuente, al pie de los mismos, en letra Arial de 8 puntos.

Instrucciones para el envío de artículos Revista ib de la Información Básica Estadística

CENTRO ANDINO DE ALTOS ESTUDIOS - CANDANE

4 ■ Notas

Los autores se hacen responsables respecto a la propiedad intelectual, por el uso de contenidos, figuras, tablas, fotografías, etc.

El artículo enviado no debe estar postulado para publicación simultáneamente en otras revistas u órganos editoriales

Cuando el documento es aceptado para su publicación, los derechos de copia del mismo pasan a ser propiedad del DANE.

El contenido de cada artículo es entera responsabilidad del autor y no compromete al Departamento Administrativo Nacional de Estadística (DANE) ni a la Revista ib.

Los artículos enviados no deben estar postulados para publicación simultáneamente en otras revistas u órganos editoriales

Para enviar artículos a la revista u obtener más información dirijase a:
cavendanop@dane.gov.co

Para obtener información sobre CANDANE y conocer el resto de nuestras publicaciones y la programación docente, visite
<http://www.dane.gov.co/candane>

Fotografía por nzbuu
Publicada originalmente en:
<http://www.flickr.com/photos/nzbuu/3187354755/>
Bajo licencia Creative Commons Atribución-No Comercial-Compartir bajo la Misma Licencia



El DANE celebra en este 2013 los primeros 60 años de su creación como Departamento Administrativo.

A través de esta publicación, CANDANE, el centro de formación y capacitación de la entidad, demuestra como en estas seis décadas nuestra institución ha ido mucho más allá de la producción de estadísticas y se consolida como una organización fuerte en el análisis de los datos, con reconocida solidez académica e investigativa.
