

ISSN 1390-6208
e-ISSN 1390-7867
Número 1 - 2013
enero - junio

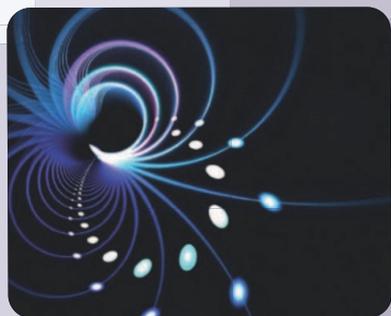
Analítica

volumen

Revista de Análisis Estadístico
Journal of Statistical Analysis



instituto nacional de estadística y censos



“Contexto”

Interpretación funcional de la lógica significativa en el desarrollo de la ciencia.

Analítica

volumen

5

Revista de Análisis Estadístico
Journal of Statistical Analysis

ISSN 1390-6208

e-ISSN 1390-7867

Analítica ★ **Volumen 5** ★ **Número 1** ★ **enero - junio 2013** ★ **Año 3** ★ **Quito - Ecuador**. Revista semestral de análisis estadístico del Instituto Nacional de Estadística y Censos, Ecuador. Publicación internacional bilingüe dedicada a estudios relacionados con ciencias sociales, económicas, análisis estadístico y de datos. Las ideas y opiniones expresadas en las colaboraciones, son de exclusividad de las autoras y los autores.

DIRECTORIO REVISTA

Roberto Castillo M.Sc.

**DIRECTOR DE ESTUDIOS ANALÍTICOS
ESTADÍSTICOS**

Paúl Medina Ph.D.

EDITOR - DIRECTOR CIENTÍFICO DE LA REVISTA

Francisco Venegas Martínez Ph.D.

Instituto Politécnico Nacional - México

Diego Chamorro Ph.D.

Université d'Evry Val d'Essonne - Francia

Silviu Glavan Ph.D.

Universidad Navarra - España

Adriana Uquillas Ph.D.

Banco Itau - Unibanco - Brazil

Carlos Machicado Ph.D.

Instituto de Estudios Avanzados en Desarrollo - Bolivia

Carlos Almeida Ph.D.

Université Catholique de Louvain - Bélgica

Ana Molina M.Sc., Ph.D.(c)

Instituto Universitario de Investigación José Ortega y

Gasset - España

COMITÉ EDITORIAL INTERNACIONAL

Borys Alvarez Ph.D.

Universidad Central del Ecuador

Petronio Alvarez Ph.D.

Universidad Central del Ecuador

Juan Mayorga-Zambrano Ph.D.

Universidad Tecnológica Israel

Wilson Pérez Ph.D.

Facultad Latinoamericana de Ciencias Sociales

Juan García M.Sc.

Universidad de Las Américas

Sergio Castillo M.Sc.

Escuela Politécnica del Ejército

COMITÉ EDITORIAL NACIONAL - ECUADOR

Lilia Quituisaca Samaniego

COORDINACIÓN EDITORIAL

Giovanny Quimbiulco

Andrés Merino (Edición en \LaTeX)

APOYO EDITORIAL

Martha Rodríguez

REVISIÓN DE ESTILO Y REDACCIÓN

Byron Sosa

DISEÑO Y DIAGRAMACIÓN PORTADA Y SEPARATAS

Indización y Registro. Analítica está indizada y/o registrada en los catálogos: COPE • Latindex • DIALNET • EBSCO • CENGAGE LEARNING • Pubindex • MIAR • RESEARCH BIBLE • DOAJ • COPE • OJS.es • ResearcherID • Mendeley • Zotero • ORCID • THOMSON REUTERS COMMUNITY • World cat • Academia.edu • ResearchGate • Google Académico • CEDIA • Universia • DSpace 3.0 • REBIUN-CRUE • AbSys7 WebOpac • Documat-Unirioja • Centro de información bibliotecario - ESPOL • Biblioteca de ingeniería eléctrica y electrónica - EPN • Biblioteca de la Universidad Guayaquil • Scoop.it • Scribd • e-bookbrowse • docstoc.

Copyright. Analítica. Se permite la reproducción total o parcial de esta revista citando la fuente.

Edición. Presentamos una re-edición de la revista Analítica volumen 5. Septiembre 2013.

El sistema tipográfico utilizado para componer esta revista fue \LaTeX , el cual es un sistema de composición orientado especialmente a la creación de libros y documentos científicos de alta calidad sobre software libre.

PRESENTACIÓN

Estimados lectores,

Es un placer presentar el quinto volumen de la Revista Analítica, en el mismo se recogen los trabajos de investigadores nacionales y extranjeros en los ámbitos económico, financiero, estadístico y sociológico. En particular se muestran los siguientes estudios:

- “Teoría y Práctica de la Discriminación en el Mercado Laboral Ecuatoriano (2007-2012)”. En este estudio se analiza la discriminación por gusto (*taste-based*), basada en la discriminación estadística (*statistical-discrimination*); evaluando las metodologías y resultados de estudios de brechas salariales y discriminación realizados sobre Ecuador.
- “Factores determinantes de la migración de los ecuatorianos”. En este estudio se evidencia el proceso migratorio, generado en el Ecuador desde inicios del actual milenio, poniendo énfasis en las características de los migrantes ecuatorianos, a partir de las semejanzas existentes entre las condiciones sociodemográficas de los hogares a los que pertenecían antes de salir del país.
- “Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación Monte Carlo”. En este estudio se estima la dotación de capital por Riesgo Operacional, utilizando modelos de pérdidas agregadas. Incluye, además, los resultados de varios experimentos de casos simulados y reales, bajo distintas funciones de distribución clásica.
- “Cuantificación de las pérdidas inesperadas ocasionadas por la delincuencia en Ecuador”. En esta investigación se estudia el costo económico que representaría para la sociedad la actividad delictiva, considerando delitos como: robo total de vehículo, robo a accesorios de vehículo, robo de vivienda, robo a personas, estafa, intimidación o amenaza y heridas.

Cabe destacar que la investigación “Cuantificación de las pérdidas ocasionadas por la delincuencia en Ecuador” ha sido expuesta en el “I Congreso Internacional de Estadística Aplicada” realizado en Ecuador y organizado por la Escuela Superior Politécnica del Chimborazo (ESPOCH); en las “Tertulias Estadísticas” realizadas por el INEC y de manera particular queremos señalar que este trabajo ha sido aceptado, luego de ser revisado por el comité científico, para su ponencia y discusión, en el *59th World Statistics Congress (WSC)*, organizado por the *Census and Statistics Department (C&SD)*, the official statistical agency of the *Hong Kong Special Administrative Region (HKSAR)*, con el aval académico y científico del *International Statistical Institute (ISI)*.

Es un hito para la Institución y el país que este trabajo, realizado en su totalidad por investigadores ecuatorianos, dos de las autoras funcionarias del Instituto y el otro autor profesor –investigador de la Escuela Politécnica del Ejército– haya sido aceptado en tan prestigioso evento científico, considerando que es el único, hasta donde conocemos de todo el Ecuador en ser aceptado.

Con agrado queremos anunciar que además de estar indexados en Latindex y Dialnet, la revista ha sido ingresada en EBSCO Publishing database y CENGAGE LEARNING, bases de datos de reconocido prestigio mundial. Esta inclusión no hace más que reconocer y valorar, una vez más, el trabajo serio y de alta calidad que se viene realizando en la edición de la revista, pues para la inclusión en las bases de datos señaladas, la revisión y cumplimientos de parámetros, altamente exigentes, han sido cumplidos en su totalidad. Debemos señalar que la revista continúa con la postulación en otras bases de datos científicas, cada vez más exigentes, todo ello con el fin de mejorar y posicionales como una de las mejores revistas del país y la región.

Finalmente, queremos reiterar nuestro compromiso de difundir los trabajos de investigadores nacionales y extranjeros, de alta calidad. Es claro que hasta el momento, casi todos los temas publicados en la revista han abordado problemas sociales, demográficos, financieros y económicos del país, a más de tener un alto valor científico, han pretendido mostrar posibles causas y soluciones, con el objetivo de que cada día nos desarrollemos en una sociedad basada en el conocimiento.

José Rosero
Director Ejecutivo

CONTENIDO

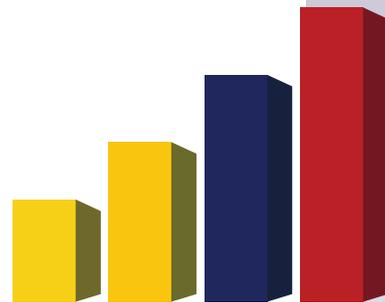
Presentación	1
Teoría y Práctica de la Discriminación en el Mercado Laboral Ecuatoriano (2007-2012)	7
Theory and Practice of Discrimination in the Ecuadorian Labor Market (2007-2012) <i>Jairo Rivera</i>	
Factores determinantes de la migración de los ecuatorianos	25
Determinants of migration of Ecuadorians <i>Patricia Cortez y Paúl Medina</i>	
Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación-Monte Carlo	39
Operational Risk Quantification using aggregate loss models and Monte Carlo simulation <i>Marco Flores</i>	
Cuantificación de las pérdidas inesperadas ocasionadas por la delincuencia en Ecuador	51
Quantification of losses caused by delinquency in Ecuador <i>Yannira Chávez, Patricia Cortez y Paúl Medina</i>	
Normas para la presentación de trabajos	65
Norms for submission of papers	68
Código de ética	71

Analítica

Teoría y Práctica de la Discriminación en el Mercado
Laboral Ecuatoriano (2007-2012)

Theory and Practice of Discrimination in the Ecuadorian
Labor Market (2007-2012)

Jairo Rivera



www.inec.gob.ec | www.inec.gob.ec/analitika

Teoría y Práctica de la Discriminación en el Mercado Laboral Ecuatoriano (2007-2012)

Theory and Practice of Discrimination in the Ecuadorian Labor Market (2007-2012)

Jairo Rivera

Pontificia Universidad Católica de Chile, Santiago, Chile

jirivera3@uc.cl

Recibido: 14 de marzo de 2013

Aceptado: 17 de mayo de 2013

Resumen

Este trabajo analiza en detalle las dos principales teorías de discriminación que han sido usadas ampliamente por los economistas: la discriminación por gusto (*taste-based*) y la discriminación estadística (*statistical-discrimination*). Además, evalúa las metodologías y resultados de los pocos estudios de brechas salariales y discriminación realizados sobre Ecuador. Posteriormente, se determina la existencia de discriminación en el mercado laboral ecuatoriano para los periodos de 2007 y 2012. En la estimación, se usa una ecuación semi-logarítmica, se realiza la corrección del sesgo de selección empleando el método de Heckman en dos etapas, y se usa la descomposición de Oaxaca-Blinder para determinar las diferencias salariales atribuibles a factores observables y no observables. Los resultados indican que las brechas salariales entre hombres y mujeres se han reducido de 15.1 %, para el año 2007, a 10.1 %, para el año 2012. Asimismo, las brechas salariales hacia grupos minoritarios han disminuido de 6.0 % a 3.7 %. Finalmente, los resultados muestran que las diferencias salariales se deben a la presencia de discriminación en el mercado laboral ecuatoriano.

Palabras clave: Brecha Salarial, Ecuación Semi-logarítmica, Descomposición Salarial Oaxaca-Blinder.

Abstract

This paper analyzes in detail the two main theories of discrimination that have been extensively used by economists; they are discrimination by taste (*taste-based*) and statistical discrimination (*statistical-discrimination*). In addition, it evaluates the methodologies and results of the few studies of wage gaps and discrimination about Ecuador. Subsequently, it determines the existence of discrimination in the Ecuadorian labor market for the periods 2007 and 2012. In estimating is used a semi-logarithmic equation, the selection bias is corrected using the Heckman method in two stages, and is done the Oaxaca-Blinder decomposition to determine the wage gap attributable to observable and unobservable factors. The results indicate that the wage gap between men and women has been reduced from 15.1 % for 2007 to 10.1 % for 2012. Also, the wage gap for minority groups decreased from 6.0 % to 3.7 %. Finally, the results show that wage differences are due to the presence of discrimination in the Ecuadorian labor market.

Keywords: Wage Gap, Semi-logarithmic equation, Wage Decomposition Oaxaca-Blinder.

Código JEL: C51, J23, J31, J71.

1 Introducción

La Real Academia de la Lengua Española (2013) define a la discriminación como la “acción y efecto de discriminar”, y a discriminar como “dar trato de inferioridad a una persona o colectividad por motivos raciales, religiosos, políticos, entre otros”.¹ Cooter (1994) realiza una definición de la discriminación en la economía: “discriminación en la vida económica por lo general consiste en la clasificación de las personas según los rasgos en lugar de la productividad”. Aigner y Cain (1977) sustentan que la discriminación es algo inevitable, y que se produce solamente cuando grupos con la misma productividad media reciben diferente salario promedio. Las implicaciones económicas de la discriminación son la motivación del presente estudio.

Desde la economía se ha buscado estimar la existencia de discriminación en el mercado laboral, y dos teorías sobresalen; por un lado, el gusto por la discriminación –*taste based*– de Gary Becker y, por otro lado, la discriminación estadística –*statistical discrimination*– de Kenneth Arrow, Edmund Phelps, y otros autores. Becker (1957) inició el estudio formal de la discriminación con rigor científico y su mayor contribución fue “desarrollar una teoría de la motivación inmaterial de la discriminación y aplicarla cuantitativamente a la discriminación de mercado”. Por su parte, Arrow (1972) contribuyó con una “teoría que enfatiza en cómo la información, o más adecuadamente, las creencias y expectativas influyen el comportamiento económico”.

En el Ecuador existen pocos estudios sobre discriminación en el mercado laboral. Éstos se han enfocado en varios aspectos: Espinoza (2009) se enfoca en la medición de brechas de género, Córdor (2010) basa su trabajo en la medición de brechas entre las etnias del país, Carrillo (2004) trabaja en brechas intersectoriales entre el sector público y privado, Carrillo y Vasconez (2011) se centran en brechas intrasectoriales en los organismos del sector público ecuatoriano, y Figueroa y Miranda (2009) estudian sobre brechas regionales. Sin embargo, dentro de los estudios se observa un sustento teórico que no se centra en discriminación, y los resultados están dispersos.

Por lo expuesto, el objetivo del artículo se concentra en abordar de una forma amplia la literatura de discriminación, aglutinar los resultados de los estudios hechos para el Ecuador, y determinar la existencia de discriminación en el mercado laboral ecuatoriano para los periodos de

2007 y 2012. En la estimación, se usa una ecuación semi-logarítmica, se realiza la corrección del sesgo de selección empleando el método de Heckman en dos etapas, y se usa el método de Oaxaca-Blinder para determinar las diferencias salariales atribuibles a factores observables y no observables.

El artículo se estructura de la siguiente manera: en la sección 2, se presentan las principales teorías de discriminación, sus avances, y una evaluación de sus resultados; además, se agrupan las investigaciones de discriminación para el Ecuador, analizando la metodología de estudio empleada y los resultados. En la sección 3, se expone la metodología y modelos usados en la estimación de brechas salariales y discriminación. En la sección 4 se presentan los principales resultados del estudio. Finalmente, en la sección 5 se señalan las conclusiones más relevantes.

2 Teorías de discriminación

En esta sección, se analizan los principales² modelos económicos que ayudan a explicar la discriminación. Iniciamos con el modelo de gusto por discriminación (*taste based*), propuesto por Becker; luego, el modelo de discriminación estadística (*statistical-discrimination*) propuesta por Arrow, Phelps, entre otros autores; por último, se realiza una evaluación empírica de los modelos y sus posteriores desarrollos

2.1 Modelo *taste-based*

Becker (1957) estuvo convencido de que es posible desarrollar una definición no ambigua de discriminación dentro del mercado, manteniendo la esencia de lo que se conoce como discriminación; para definirla usa como herramienta de medida al dinero e introdujo un concepto denominado “gusto por la discriminación”³. De igual manera, se crea la figura de *coeficiente de discriminación (DC)*⁴ como puente entre los costos de dinero de una transacción y los costos netos de esa transacción, y se ejemplifica así:

- para los empleadores, el salario de un factor es w ; pero al existir discriminación, el salario neto de ese factor se convierte en $w(1 + d_i)$;

¹De la misma forma, se define la discriminación positiva como la “protección de carácter extraordinario que se da a un grupo históricamente discriminado, especialmente por razón de sexo, color de piel, lengua o religión, para lograr su plena integración social”. Sin embargo, normalmente la discriminación se liga a factores negativos, pudiendo inferirse su concepto tomando el opuesto de la discriminación positiva.

²La discriminación por gusto y la discriminación estadística son las principales corrientes que han usado los economistas para analizar la discriminación, y cada una ha tenido desarrollos posteriores apoyados por los avances en teoría de juegos y econometría. Bertrand et al. (2005) proponen un nuevo enfoque para analizar la discriminación utilizando factores de la psicología de las personas, y la denominan discriminación implícita. La hipótesis de que las personas discriminan sin intención ni conciencia es el fundamento de la discriminación implícita, y se plantea que una de sus ventajas es que se podría reducir la discriminación sin afectar las preferencias de los individuos.

³Becker lo define así: “si una persona tiene un gusto por la discriminación debe actuar como si fuera a pagar algo, ya sea directa o en una reducción de los ingresos, que se asocia con algunas personas en contra de otras. Cuando la discriminación ocurre, la persona debe, de hecho, pagar o perder ingresos por ese privilegio”.

⁴Jorgensen y Fenger (2009) definen adecuadamente el DC desarrollado por Becker, así: “el coeficiente de discriminación es el equivalente en dinero personal de los costos no monetarios asociados con la desutilidad de estar afiliado a una persona u objeto”.

- para un empleado con salario w_j , el trabajar con el factor discriminado convierte su salario neto en $w_j(1 - d_j)$, siendo d_j el DC;
- para un consumidor que va a comprar un bien que tiene costo p producido por el factor discriminado, el costo neto se convierte en $p(1 + d_k)$, siendo d_k el DC.

El DC es definido como un número natural (cero o positivo), y se usa como una representación cuantitativa del gusto por discriminación, que es la base para cuantificar empíricamente la importancia de la discriminación. En el presente estudio, abarcaremos la teoría concerniente a la discriminación por parte de los empleadores. Becker considera que la discriminación de mercado está ligada netamente con la productividad, y define la discriminación por parte de los empleadores como "el rechazo a contratar una persona con un valor de producto marginal más alto que su costo marginal".

Becker usa los supuestos de que el mercado se maneja en competencia perfecta, la producción tiene rendimientos constantes a escala, y los trabajadores con color de piel blanca y negra son sustitutos perfectos en la producción. La función de utilidad del empleador depende del beneficio y del número de trabajadores que contrata, hacia los cuales tiene gusto por discriminación. Además, los empleadores pueden o no tener prejuicios raciales que los lleven a discriminar, y se supone que son de piel blanca, mientras que los trabajadores pueden ser blancos o negros, y el grupo discriminado es el de color negro. Tomaremos la notación de Charles y Guryan (2008); con ello, la función de utilidad del empleador se define así:

$$U_i = \pi_i - d_i l_n$$

donde π_i representa los beneficios del empleador; así,

$$\pi_i = f(l_b + l_n) - w_b l_b - w_n l_n$$

ya, w_b representa el salario de las personas de color blanco, w_n el salario de color negro, $f(\cdot)$ es la función de producción con retornos constantes a escala, y el empleador escoge entre $(l_b + l_n)$ para maximizar su utilidad. Asimismo, la maximización de utilidad cumple con dos condiciones, que sintetizan la esencia del modelo de Becker:

$$f'(l_b + l_n) - w_b \leq 0 \text{ con igualdad si } l_b > 0$$

$$f'(l_b + l_n) - w_n - d_i \leq 0 \text{ con igualdad si } l_n > 0$$

donde dichas condiciones representan que el empleador contratará trabajadores hasta el punto en que su producto marginal sea igual a su costo marginal. Para el trabajador sin discriminación, el costo marginal es su salario; pero para el trabajador discriminado, su costo es más alto, ya que incorpora su salario y los prejuicios del empleador. Con ello, los trabajadores elegirán contratar trabajadores negros solo si:

$$w_n + d_i \leq w_b$$

En el corto plazo, los mercados se deben aclarar para que exista equilibrio. Es aquí donde aparece el discriminador marginal d_i^* como elemento fundamental para determinar si una persona contrata o no a un trabajador. Las personas con menor prejuicio que d_i^* contratarán trabajadores negros, y aquellos con menor prejuicio contratarán solo blancos; con ello, el equilibrio es:

$$w_n^* + d_i^* = w_b^*$$

El proceso usado por Becker para desarrollar su concepto de discriminación toma en cuenta variables importantes, tales como: i) distancia social y física entre un individuo que actúa como discriminador y el estatus socioeconómico relativo del grupo del discriminado, ii) si trabajan en producción, interesa la sustituibilidad del factor, iii) el número relativo del grupo discriminado dentro de la sociedad, iv) cantidad de conocimiento acerca del grupo discriminado, v) ubicación geográfica y cronológica, y vi) diferencias en personalidad. Becker llega a la conclusión de que la discriminación afecta negativamente al ingreso del grupo discriminado, pero también al propio ingreso del grupo discriminador tomando como ejemplo el comercio internacional. Además, si el grupo discriminado reacciona con represalias hacia el grupo discriminador, el salario del primer grupo disminuye aún más. Con ello, concluye que la discriminación afecta más a las minorías que a las mayorías.

Finalmente, Becker (1957) analiza la discriminación de mercado por parte de los empleadores. Obtiene las siguientes tres relaciones: i) la distribución del gusto por la discriminación en los empleadores incrementa el empleo del grupo no discriminado en detrimento de la contratación del grupo discriminado, si el empleador es muy proclive a discriminar; sin embargo, esta decisión puede ser contraproducente, ya que se podría perder de un buen empleado con productividad marginal alta que pertenece al grupo discriminado, ii) la cantidad de competencia relativa al monopolio es un factor que afecta a la discriminación ya que, a mayor competencia en la industria o mercado existe menor margen para la discriminación; los empleadores que discriminan deberán incurrir en mayores costos, lo cual puede ocasionar que queden fuera del mercado; y, iii) el número relativo de trabajadores del grupo discriminado es un factor determinante en la discriminación por parte de los empleadores; así, a medida que se incrementa el grupo de trabajadores discriminados, se produce un aumento de la discriminación hacia ese grupo y sus salarios se reducen.

2.2 Modelo *statistical-discrimination*

La discriminación estadística nace ante el auge de investigaciones sobre "información limitada" a inicios de los años setenta. Entre sus principales expositores se destacan Kenneth Arrow (1972), Edmund Phelps (1972), y Aigner y Cain (1977).

Arrow (1972) enfatiza que la diferencia entre los salarios de las personas se debe a que dentro del mercado exis-

ten grupos que tienen sus características propias, y el mercado valora dichas características. Es decir, la noción de discriminación estadística⁵ implica que el mercado otorga valor no solo a la productividad, sino también a las características personales. Describe su estudio como una expansión del modelo previo de Becker, pero ligándolo más hacia la teoría de equilibrio general, pensado más como un contraste y menos como un acuerdo.

El análisis realizado por Arrow (1973) sobre el trabajo de Becker, donde se incorpora el gusto por discriminación dentro de la función de utilidad generando que las empresas que tienen mayor discriminación obtengan menores beneficios económicos, considera que la estructura de gustos por discriminación da una interesante explicación en el corto plazo pero que, a largo plazo, parece no resistir las presiones competitivas. Ante ello, escribe su famosa frase “*el modelo predice la ausencia del fenómeno para el cual fue desarrollado*”. Asimismo, sostiene que la relación entre el prejuicio y los salarios, planteada por Becker, puede sobrevivir en el largo plazo solo en presencia de tres elementos: si el mercado no es perfectamente competitivo, si existen importantes no convexidades, o si hay problemas de información.

Además, Arrow (1998) critica duramente la forma en que es tratada la maximización de utilidades en el modelo *taste-based*. La inclusión de nuevas variables dentro del problema del productor, como la utilidad negativa de contratar trabajadores discriminados, genera desafíos para explicar la teoría de elección racional. Esto produce que la explicación del modelo pueda caer en tautología, y que la medición de la función de utilidad sea casi imposible de obtener para empresas grandes.

Phelps (1972), por su parte, realiza una distinción entre su teoría naciente y la de Becker, argumentando que la discriminación estadística puede ser aplicada a todo tipo de empleador o trabajador; se basa en que no es necesaria la existencia de un gusto por discriminación hacia contratar y trabajar junto a trabajadores negros o mujeres, a diferencia del *taste-based*, que necesita la presencia del gusto por discriminación. La importancia de la información estadística es clave para el modelo de Phelps, y se usa para suplir o reducir la información imperfecta del mercado.

Debido a la información imperfecta, el empleador no observa el nivel de habilidad de los solicitantes de trabajo, pero sí observa el grupo al que pertenecen $j \in (b, n)$, donde b es blanco y n es negro. Tomamos la notación de Fang y Moro (2011) para describir el modelo de Phelps.

Las habilidades de los trabajadores se denotan con q , y se asume que son iguales a su producto marginal; además, tiene una distribución normal, así $N(\mu_j, \sigma_j^2)$. Los trabajado-

res observan el grupo al que pertenece el trabajador, y una señal con ruido de productividad, así $\theta = q + \varepsilon$, donde ε se distribuye normalmente, así $N(0, \sigma_{\varepsilon j}^2)$. Cada empleador deduce de θ el valor esperado de q con la información disponible, incluyendo la identidad de grupo. La habilidad y la señal se distribuyen conjuntamente con una distribución normal, y la distribución condicional de q dado θ es normal, con media igual a un promedio ponderado de la señal y la media del grupo incondicional. Así,

$$E(q|\theta) = \frac{\sigma_j^2}{\sigma_j^2 + \sigma_{\varepsilon j}^2} \theta + \frac{\sigma_{\varepsilon j}^2}{\sigma_j^2 + \sigma_{\varepsilon j}^2} \mu_j$$

Como resultado se obtienen las condiciones en las cuales el salario tenderá a ser más cercano a la productividad o más cercano a la media de la población. Si la señal es muy ruidosa (varianza de ε alta), el valor esperado condicional de la productividad de los trabajadores se encuentra cerca de la media de la población, independientemente del valor de la señal; y si la señal es muy precisa (varianza de ε pequeña), entonces la señal proporciona una estimación precisa de la capacidad del trabajador. En su análisis, Phelps propone dos casos de desigualdad:

Caso 1: las señales de grupos son igualmente informativas, pero un grupo tiene mayor inversión promedio de capital humano y ($\sigma_{\varepsilon b} = \sigma_{\varepsilon n} = \sigma_{\varepsilon}$), ($\sigma_b = \sigma_n = \sigma$) y ($\mu_b > \mu_n$). Los empleadores actúan racionalmente y otorgan menores salarios a los trabajadores del grupo n , ya que generan una menor productividad esperada, a pesar de que existe la reciben la misma señal del grupo.

Caso 2: las señales de los grupos son diferentes ($\sigma_{\varepsilon n} > \sigma_{\varepsilon b}$), mientras que las distribuciones incondicionales de habilidades son las mismas entre los dos grupos ($\sigma_b = \sigma_n = \sigma$) y ($\mu_b = \mu_n = \mu$). Con ello, los empleadores otorgan menores salarios a los trabajadores del grupo con mayor señal. Este caso es una muestra de discriminación debido a la información que llega al empleador, y ejemplifica lo que sucede en el mercado laboral.

La frase final del estudio de Phelps es: “*la discriminación no es menos perjudicial para sus víctimas por ser estadística, ni menos importante para ser contrarrestada por la política social*”.

2.3 Evaluación empírica de los modelos

Como defensores del modelo *taste-based* aparecen Charles y Guryan (2008), y realizan un trabajo en donde se evalúan y comprueban los principales postulados del modelo de Becker para diferentes Estados miembros de Estados Unidos. Primeramente, se muestra que los salarios de la

⁵Raphael (2002) realiza una definición interesante de la discriminación estadística, así: “cuando la información descrita por el color de piel es relevante, los agentes atribuyen las características prescritas del grupo al sujeto individual, por lo que se genera la construcción de una ‘identidad social virtual’ para el sujeto que puede no tener relación con las características reales de la persona”. Por su parte, Cahuc y Zylberberg (2004) consideran que la discriminación estadística estudia cómo individuos con idénticas habilidades pero pertenecientes a diferentes grupos no tienen senderos de carreras equivalentes, debido a la calidad promedio, real o imaginaria, del grupo al cual pertenecen. Por otro lado, Schwab (1999) considera que los modelos de discriminación estadística no asumen prejuicios o gustos por parte de los empleadores para trabajar con determinada fuerza laboral, sino más bien se dedican al estudio de las características grupales para predecir atributos individuales en un mundo de información limitada; finalmente, denomina en forma acertada, como discriminación estadística, a aquella que se deriva de las correlaciones estadísticas.

gente de piel negra en comparación con los de la gente de piel blanca tienen tres comportamientos: i) varían negativamente ante el nivel de prejuicio de la persona "marginal" en un Estado, ii) varían negativamente con el prejuicio en el extremo inferior de la distribución de los prejuicios, pero no se ven afectados por los prejuicios de las personas más perjudicadas en un Estado, y iii) varían negativamente con la fracción de un Estado que es negro. Adicionalmente, se presentan pruebas para demostrar que las brechas salariales por color de piel son más grandes cuanto más integrada racialmente está la mano de obra del Estado. Además, se indica que las predicciones principales del modelo de Becker, mediante demostración teórica, sobreviven a los efectos de la competencia a largo plazo. Asimismo, se considera que la discriminación salarial sería menor si el mercado tuviera la capacidad de segregar; pero ante mercados más amplios y mayor población, los objetivos de segregación no se alcanzan. Finalmente, se demuestra que los lugares donde las personas de color negro reciben salarios más bajos es donde las personas de piel blanca tienen una especie de gusto por discriminación mayor, ya que existe mayor proporción de blancos que se oponen a los matrimonios interraciales, o blancos que no votarían por un presidente negro, y usan estas variables como proxys del gusto por la discriminación.

Una modificación interesante al modelo *taste-based* se presenta en McAdams (1995), donde se incorpora el hecho de que a los blancos no les cuesta asociarse con los negros siempre y cuando la relación mantenga jerarquías; pone como ejemplo a la práctica común de los blancos que contratan a trabajadores negros para las labores domésticas en sus hogares. La psicología forma parte de este modelo, ya que se presume que los blancos forman un grupo socialmente conectado que invierte en la elevación de su autoestima mediante la subordinación de los negros. Además, se explica el por qué los trabajadores de color de piel negra no ocupan puestos de rango superior en las instituciones.

Mediante el uso de modelos de búsqueda, Black (1995) muestra que las predicciones de Becker para el largo plazo pueden persistir en un entorno competitivo, con la particularidad de que el acceso a información sea costoso y no exista información completa. En armonía con lo anterior, Lang y Lehman (2010) afirman que los modelos *taste-based* pueden generar diferenciales de salarios y desempleo cuando se combinan con cualquier modelo de búsqueda sea aleatoria o dirigida. Con ello, se realza el uso de modelos de búsqueda con el *taste-based*.

Un estudio llamativo que aplica el *taste-based* fue el desarrollado por Levitt (2004). En la investigación se trabaja con datos del programa "The Weakest Link" y se encuentra que los participantes tienen una tendencia a votar por las personas que representan a grupos minoritarios en las primeras rondas del juego. Sin embargo, posteriormente, con la información obtenida durante el juego, los jugadores cambian de estrategia y trasladan su votación hacia

aquellos que no les agradan, obviamente tomando en cuenta más variables. El *taste-based* de los participantes los lleva a discriminar, en su mayoría, en contra las personas de la tercera edad.

Sin embargo, entre las críticas principales al modelo del *taste-based* se encuentra la dificultad de medirlo. Ante ello los modelos *statistical-discrimination* han tomado mayor fuerza y desarrollo.

Entre los estudios a favor del *statistical-discrimination* se destaca Loury (1998), quien realiza una distinción entre las teorías basadas en el *taste-based* de Becker y la *statistical-discrimination* de Arrow y Phelps. Loury considera que los modelos basados en la información estadística realizan una descripción más realista del mercado laboral y la discriminación hacia los negros. El uso del color de piel como factor de discriminación es una cuestión de fácil observación, que se correlaciona con algunas otras que son difíciles de observar. Sin embargo, a su vez, Loury señala que la discriminación estadística debe ser tratada con "sospecha legal" ya que si no es bien analizada, puede generar problemas en la eficiencia de las asignaciones de mercado y en la equidad.

Arrow (1998) destaca un aspecto interesante, que se comprueba en varios estudios empíricos, y es que la diferencia de salarios entre hombres y mujeres en la historia moderna de Estados Unidos es mínima, si ambos realizan el mismo trabajo. Eso, a primera vista, nos daría luces de que no existe problema de discriminación; sin embargo, lo que sucede en realidad es que aparece la segregación laboral, ya que las mujeres no pueden acceder a ciertos puestos de trabajo; este caso es comparable a la segregación residencial.

List (2004) realiza una atractiva aplicación de la discriminación estadística, al investigar la compra y venta de las tarjetas de deportes. Los grupos minoritarios reciben menores ofertas de precios por la reserva de las tarjetas y, pueden llegar a obtener el mismo precio que se pagaría a un grupo normal solamente después de perder recursos como tiempo. Adicionalmente, se encuentra que las personas que realizan transacciones con más experiencia son aquellas que en la práctica discriminan más, y la causa debe ser, en sí misma, su historial de transacciones, es decir, su estadística.

Uno de los problemas de la discriminación estadística es que trabaja con promedios de grupos y, con ello, el ser individual se ve supeditado a las características de su grupo. La asignación eficiente de recursos en economía se produce cuando existen valoraciones marginales; sin ellas, es probable que se cometan errores. Además, el problema puede agravarse, como en el modelo de Aigner y Cain (1977), en donde los incentivos individuales para entrenarse hacia el grupo discriminado son pocos porque, a pesar de que se eduquen, seguirán percibiendo una parte de su remuneración de acuerdo a los salarios promedios del grupo al que pertenecen. De la misma forma, Lundberg y Startz (1983) analizan cómo la discriminación estadística puede agravar

las distorsiones de mercado ya que las personas del grupo discriminador tienen más incentivos para invertir en formación, en comparación con los discriminados. Coate y Loury (1993) denominan a este proceso el apareamiento de "estereotipos negativos" hacia el grupo discriminado, pudiendo perpetuar el efecto en los salarios, equidad y eficiencia en el mercado. Cahuc y Zylberberg (2004) resumen acertadamente el problema, señalando que la discriminación estadística puede generar una inequidad persistente entre grupos, ya que las decisiones de los empleadores influyen en la educación de los grupos discriminados; con ello, un grupo que no observa retornos salariales mayores ante mejor educación, no tiene los incentivos necesarios para educarse; esto puede afectar tanto su instrucción presente como la educación que brindarán a sus hijos, lo cual perpetuará la discriminación. En la misma línea, Schwab (1986) estudia cómo la discriminación estadística puede agravar las diferencias de la oferta de trabajo, porque los trabajadores encasillarán su búsqueda solo hacia ciertos empleos, y los grupos discriminados tenderán hacia los que ofrecen menor salario.

Por otro lado, Sunstein (1991) ha encontrado evidencia para contradecir el modelo de Lundberg y Startz; plantea que, debido a que los retornos a la educación en la práctica son mayores para la gente con color negro de piel que para los blancos, eso puede estimular la inversión en formación tanto para el empleador como para el trabajador en favor de dicho grupo discriminado. Los efectos de desincentivo hacia los grupos minoritarios se transforman hacia una mejora de la equidad social.

Finalmente, en la parte metodológica han existido aportes interesantes al trabajar con el modelo de discriminación estadística; se los detallará a continuación:

Heckman (1979) analiza el problema de sesgo por auto-selección. Primero, indica que el sesgo de selección muestral puede surgir en la práctica por dos razones: puede existir auto-selección por los individuos, o por decisiones de selección de muestras por procesadores o analistas de datos. Plantea, además, que el sesgo de auto-selección se sustenta en el problema de variables omitidas, y eso se debe a que las funciones de regresión ajustadas confunden a los parámetros de comportamiento con los parámetros de la función de determinación de la probabilidad de entrada en la muestra. Finalmente, mediante la incorporación de un corrector de sesgo se pueden corregir la matriz de varianza-covarianza y los errores estándar.

En el tema de descomposición salarial, Cahuc y Zylberberg (2004) describen los aportes de Oaxaca y Blinder en los años setenta. Las técnicas son muy similares, y consisten en comparar los salarios de un grupo con los del grupo discriminado. Mediante la estimación de las ecuaciones de salario separadamente para ambos grupos, se obtienen valores referenciales que se introducen en la metodología de descomposición. Como resultado del modelo, se puede distinguir entre los factores observables que afectan a la brecha

salarial y los no observables. Los factores no observables se atribuyen a discriminación. Como lo señalan O'Donnell et al. (2008), existen varias descomposiciones salariales derivadas de la inicial de Oaxaca-Blinder, aunque la esencia de factores explicados y no explicados por el modelo se mantiene.

Heckman (1998) realiza una crítica al modelo de discriminación estadística en estudios contemporáneos sobre discriminación. Observa que los autores confunden discriminación de la empresa con discriminación de mercado, y generalizan los resultados sin mayor cuidado. Además, indica que es el margen lo que debe interesar a los economistas al estudiar la discriminación. Finalmente, asegura que el modelo de Becker está mal interpretado cuando se asevera que en el largo plazo la discriminación desaparecerá; plantea que, por el contrario, más bien se reducirá, tal como está sucediendo en el mercado laboral de Estados Unidos.

2.4 Caso: Ecuador

A nivel mundial existe una vasta literatura sobre la medición de la discriminación salarial usando como partida el retorno de la educación y, en esencia, el modelo Minceriano. Mincer (1974) determinó que las variables relevantes para determinar el retorno de la educación son: nivel de educación, experiencia y experiencia al cuadrado.

En el Ecuador, sin embargo, existe un número reducido de estudios que se dedican a explicar la discriminación, los retornos educacionales y las brechas salariales. Los estudios existentes se han centrado en la medición de brechas en las siguientes categorías: género, etnia, intersectoriales, intrasectoriales y regionales.

Dentro de los estudios intersectoriales sobresale Carrillo (2004), quien elabora un análisis de regresión tradicional (MCO) que le permite identificar el promedio de las diferencias salariales entre el sector público y privado; utiliza la ecuación semi-logarítmica minceriana para explicar los determinantes del salario de la siguiente manera:

$$\ln(w_i) = X_i\beta + \delta * P_i + \varepsilon_i$$

donde w_i corresponde al salario por hora del asalariado i , X_i corresponde a un vector con variables explicativas que determinan el nivel del salario, β es un vector de parámetros, P_i es una variable *dummy* que toma el valor de uno si el individuo trabaja en el sector público, δ es un coeficiente escalar, y ε_i es una variable aleatoria que incluye todos los otros factores que forman parte del salario y que no son explicados por las variables independientes del modelo.

Carrillo obtuvo como principal conclusión que los trabajadores del sector público obtienen en promedio 18% más que sus colegas del sector privado. Además, señaló que existen diferencias salariales que dependen del nivel de salario, en donde, las brechas son más marcadas entre los que menos ganan y son casi nulas entre los que más ganan.

Por otro lado, en los estudios de brechas de género se destaca Espinoza (2009), con un trabajo sobre las diferencias salariales entre hombres y mujeres en el Ecuador. Aplica en su trabajo un modelo de dos etapas de Heckman, que introduce a la función de ingresos de Mincer una variable que corrige el sesgo de selección. Esta variable estima la probabilidad de que un individuo participe en el mercado laboral. La primera etapa corresponde un modelo probabilístico, donde se construye la variable lambda (λ), que es la probabilidad ajustada de que el individuo i participe en el mercado laboral. La segunda etapa corresponde a la estimación de la función de ingresos por MCO, incluyendo como variable explicativa a λ , en el modelo de Mincer. Entonces, utiliza el siguiente modelo:

$$\ln(w_i) = \beta Y_i + \gamma \lambda + \mu_i$$

Las conclusiones primordiales a las que llega Espinoza son: los hombres ganan más que las mujeres en toda la distribución por cuantiles, y el retorno de la educación en los hombres varía entre 6.4 % y 7.4 %, mientras que para las mujeres se ubica en entre 5.9 % y 7.4 %.

Usando un análisis muy similar al anterior, Córdor (2010) realiza un trabajo sobre las diferencias en retornos salariales y brechas entre las etnias del país. Luego de realizar una ecuación de Mincer y la corrección de sesgo de selección de Heckman, se procede a realizar una descomposición salarial de Oaxaca-Ransom. Los principales resultados del estudio consistieron en encontrar la presencia de discriminación entre los indígenas y los no indígenas: los primeros perciben 49.6 % menos remuneración que los segundos. Esta tendencia de discriminación se mantiene para el período 2004-2008 que abarca el estudio.

En cuanto a estudios intrasectoriales, para comparar los niveles de salarios entre las instituciones públicas, Carrillo y Vásquez (2011) usan el método de Mínimos Cuadrados Ordinarios (MCO), con incorporación de variables *dummies*:

$$\ln(w_i) = X_i\beta + S_{pi}\delta_p + \Gamma_i + \varepsilon_i$$

donde w_i es el logaritmo del ingreso anual del empleado i , X_i es un vector de características productivas de cada tra-

bajador i , β es el retorno de las características X_i , S_{pi} es la variable *dummy* de la institución p de cada trabajador público i , Γ_i es una *dummy* que controla la diferencia entre los dos años de estudio (0 es 2008 y 1 es 2009), y ε_i es el término de error. Los resultados indican que las instituciones que tienen altas responsabilidades y poder político reciben un mayor salario (de hasta más de 200 %); además, al hacer un análisis por quintiles, se concluye que existen brechas salariales entre sectores y dentro de las instituciones.

Figueroa y Miranda (2009) proponen una función de salario ampliada, que sigue la siguiente fórmula funcional:

$$\ln(Y_i) = \alpha + \beta_1 PR_i + \beta_2 SE_i + \beta_3 UN_i + \gamma_1 E_i + \gamma_2 E_i^2 + \Gamma_1 G_i + \varphi_1 A_i + \varepsilon_i$$

donde E_i son los años de experiencia, E_i^2 son los años de experiencia al cuadrado, PR_i son los años de educación primaria, SE_i son los años de educación secundaria, UN_i son los años de educación universitaria, G_i es una variable *dummy* sobre la región donde se encuentra, A_i es una *dummy* sobre el área geográfica, y ε_i es el término de error.

Las conclusiones de Figueroa y Miranda muestran que no se pudo comprobar completamente su hipótesis del retorno de inversión por región y área geográfica, debido a que el número de personas que han adquirido educación superior en el área rural es demasiado pequeño y no permite estimar la tasa de retorno; además, hay diferencias en la facilidad para obtener información respecto del área urbana. Sin embargo, si logra obtener resultados para los niveles primarios y secundarios para las áreas urbana y rural. El estudio concluye que cada año de educación primaria adquirido aporta un incremento mayor en el área urbana; para la educación secundaria, el incremento es similar en ambas áreas. Por otra parte, a nivel de región, la educación superior otorga mayores retornos en el ingreso en la Costa respecto de la Sierra; en el caso de la Amazonía, el acceso a educación superior es escaso.

En la Tabla 1⁶ se presenta un resumen de los principales estudios sobre brechas salariales realizados para el Ecuador.

Tabla 1. Resumen de los principales estudios sobre brechas salariales en el Ecuador

	Carrillo	Espinoza	Córdor	C. y V.	F. y M.
Año de Publicación	2004	2009	2010	2011	2009
Fuente de Datos	EIGHU	ECV	ENEMDUR	SRI	ENEMDUR
Año de los Datos	2002	2006	2004-2009	2008-2009	2003-2008
Corrección de Sesgo	no	sí	sí	no	no
Brecha salarial	intersectorial	género	etnia	intrasectorial	regional

Fuente: Cada uno de los estudios citados. Elaboración: Autor.

⁶ECV= Encuesta de Condiciones de Vida, ENEMDUR=Encuesta de Empleo y Desempleo Urbano y Rural, SRI=Servicio de Rentas Internas.

3 Metodología y modelo

Como bien lo reconoce Becker (1993)⁷, las mediciones de discriminación basadas en el *taste-based* son deseables, pero complicadas de obtener. En ese sentido, los estudios empíricos se han centrado en el uso de *statistical-discrimination*. En el presente trabajo determinaremos la existencia de discriminación en el mercado laboral ecuatoriano usando el segundo modelo debido a la disponibilidad de información.

3.1 Datos

Los datos que se usan corresponden a la Encuesta de Empleo y Desempleo Urbano y Rural (ENEMDUR), construida por el Instituto Nacional de Estadística y Censos del Ecuador (INEC). La ENEMDUR es una encuesta a personas y hogares que se realiza mensualmente en el Ecuador. Los períodos analizados son diciembre del año 2007 y diciembre del año 2012; en ellos se encuestó a una muestra de 76.922 y 73.686 personas respectivamente.

La ventaja de trabajar con la ENEMDUR es su periodicidad; ello nos permite realizar un análisis actual de la población ecuatoriana, y comparar con lo ocurrido en el último quinquenio. Existen otras encuestas en el Ecuador, como la Encuesta de Condiciones de Vida o la Encuesta de Ingresos y Gastos Urbanos, que si bien son una fuente importante de datos para analizar a los hogares, ambas se realizan de manera esporádica.

Con la intención de homogenizar a la muestra se la ha restringido tomando criterios de edad, lo cual es ampliamente usado en la literatura. Se ha considerado a las personas que se encuentran entre los 24 y 65 años de edad. En los estudios previos para Ecuador, como en Carrillo (2004), se restringe en mayor medida a la muestra con criterios adicionales, pero ese trabajo no realiza corrección de sesgo de selección. En los demás trabajos que realizan corrección del sesgo, las restricciones son principalmente por edad.

Ahora se presenta una breve descripción de los datos de la muestra, enfocada en los insumos principales de la ecuación de Mincer⁸. En la Tabla 2 se observa que el salario promedio por hora para el año 2012 se ubica en 2.59 dólares, teniendo un crecimiento de 38.5 % con respecto al año 2007. El promedio de años de educación también ha aumentado entre períodos, aunque en menor medida (6.6 %), y para el año 2012 se ubica en 10.01. La experiencia se ha mantenido

relativamente estable y se ubica, para el año 2012, en 26.76 años promedio.

Tabla 2. Estadísticas descriptivas

Población general		
	2012	2007
salario por hora	2.59	1.87
años de educación	10.01	9.39
experiencia	26.76	26.22

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Con el objetivo de analizar las dos fuentes de discriminación que en general estudia la literatura, discriminación por género y a grupos minoritarios, se presentan las tasas de participación en la Población Económicamente Activa (PEA), y una breve descripción de los estadísticos promedios para subgrupos.

En la Figura 1, se observa que la participación de la mujer se ha mantenido relativamente estable entre períodos, ubicándose en 40.5 % para el año 2012. Asimismo, la Tabla 3 indica que el salario promedio por hora en el año 2012 para los hombres se ubica en 2.64 dólares, mientras que el de las mujeres en 2.49 dólares. La brecha de salarios a favor de los hombres entre períodos se ha reducido de 15.2 % para el año 2007 a 6.0 % para el 2012. Por otro lado, el promedio de años de educación para los hombres en el año 2012 se ubica en 9.78 y, para las mujeres, en 10.55. La brecha de escolaridad a favor de las mujeres ha aumentado entre períodos de 4.1 % a 7.3 %. Por último, los hombres tienen más experiencia en ambos períodos, y la brecha a favor de los hombres se ha incrementado levemente.

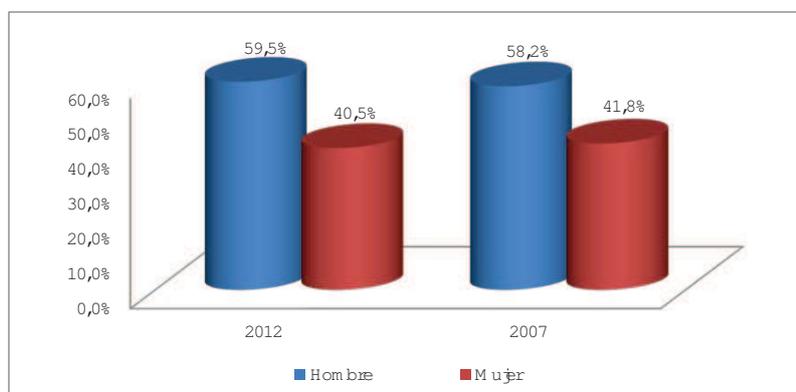
En la Figura 2, se observa un incremento de la participación de los grupos minoritarios⁹ en la PEA, llegando a ser de 15.3 % para el año 2012. De igual forma, la Tabla 4 indica que el salario promedio por hora en el año 2012 para blancos y mestizos se ubica en 2.69 dólares, mientras que el de los grupos minoritarios es de 1.99 dólares. La brecha de salarios a favor de los blancos y mestizos entre períodos se ha reducido de 50.0 % para el año 2007 a 35.2 % para el 2012. El promedio de años de educación para los blancos y mestizos en el año 2012 se ubica en 10.62, y para las mujeres en 7.21. La brecha de escolaridad a favor de los blancos y mestizos ha disminuido entre períodos de 50.5 % en el 2007 a 47.3 % en el 2012. Las personas de grupos minoritarios tienen en promedio más experiencia en ambos períodos, y la brecha a su favor ha aumentado.

⁷En palabras de Becker: "La evidencia sobre los beneficios no percibidos, los salarios y los precios no suelen estar disponibles, por lo que la discriminación en contra de un grupo se mide comparando los ingresos de los miembros del grupo con las ganancias de la 'mayoría' que tienen los mismos años de escolaridad, experiencia laboral, y otras características mensurables".

⁸La experiencia laboral se calcula en años, y corresponde al residuo entre la edad de la persona y sus años de escolaridad, y menos 6 años. Es necesario restar los 6 años debido a que esa es la edad promedio a la que se empieza a estudiar y se la cataloga como experiencia potencial.

⁹Usando como criterio los grupos de autodefinition étnica de la ENEMDUR, el autor ha definido dentro de grupos minoritarios: personas con piel de color negro, afroecuatorianos, mulatos, indígenas y montubios.

Figura 1. Participación de la mujer en la PEA



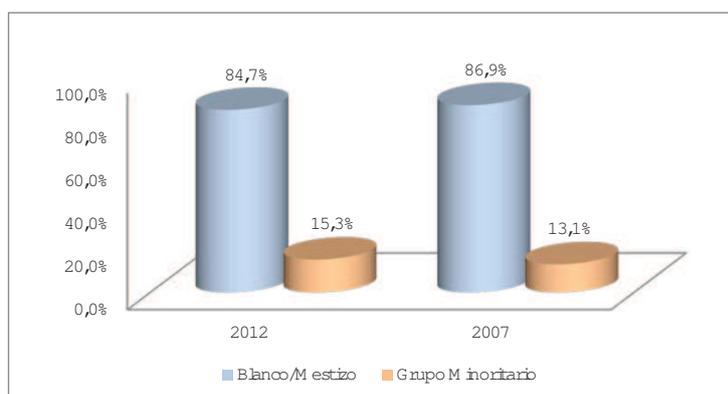
Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Tabla 3. Estadísticas descriptivas. Por Género

	2012		2007	
	Hombre	Mujer	Hombre	Mujer
salario por hora	2.64	2.49	1.97	1.71
años de educación	9.78	10.55	9.23	9.62
experiencia	27.27	25.99	26.59	25.69

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Figura 2. Participación de los grupos minoritarios en la PEA



Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Tabla 4. Estadísticas descriptivas. Por Grupo Étnico

	2012		2007	
	Mestizo/Blanco	G. Minoritario	Mestizo/Blanco	G. Minoritario
salario por hora	2.69	1.99	1.95	1.30
años de educación	10.62	7.21	9.83	6.53
experiencia	26.17	30.00	25.77	29.23

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

3.2 Modelo

En la estimación del modelo se utiliza el método de Mínimos Cuadrados Ordinarios (MCO) con el objetivo de medir el promedio de las diferencias salariales en la población. Adicionalmente, se realiza la corrección del sesgo de selección en dos etapas como lo recomienda Heckman. Además, se usa el método de Oaxaca-Blinder para determinar las diferencias salariales atribuibles a factores observables y no observables.

El modelo consiste en una ecuación semi-logarítmica con el fin de analizar los determinantes del salario:

$$\ln(w_i) = \beta_0 + x_i\gamma + t_i\theta + \varepsilon_i$$

donde C corresponde al salario por hora del asalariado i , x_i corresponde a un vector con variables explicativas que determinan el nivel del salario, γ es un vector de parámetros, t_i es un vector de variables *dummy*, θ es un vector de parámetros de las *dummies*, y ε_i es una variable aleatoria que incluye todos los otros factores que forman parte del salario y que no son explicados por las variables independientes del modelo. Los signos esperados de las variables son, de acuerdo al trabajo de Sapelli (2003), positivos para los años de educación, positivos para los años de experiencia laboral, y negativos para la experiencia al cuadrado.

La corrección del sesgo de selección se realiza estimando dos regresiones. La primera estimación se realiza a base de la ecuación de participación, entendida como la probabilidad de que el individuo participe en el mercado laboral contra los factores que pueden incidir en la elección.

$$p_i = \beta_0 + z_i\phi + u_i$$

donde p_i corresponde a la probabilidad de participar en el mercado laboral, z_i corresponde a un vector con variables explicativas que influyen en la decisión de participar, ϕ es un vector de parámetros, y u_i son los errores.

La segunda estimación es la misma ecuación de Mincer anterior, incorporando la corrección de selección realizada por λ como inversa del ratio de Mill, así:

$$\ln(w_i) = \beta_0 + \lambda x_i\gamma + \lambda t_i\theta + \varepsilon_i$$

Con ello se ha corregido el sesgo de selección. Usando el argumento de Torres y Celton (2009), la importancia del uso del λ se determina al estimar la regresión, siempre y cuando su coeficiente sea significativo. En el caso de no ser representativo, se podría trabajar directamente sin la necesidad de la corrección.

El método de descomposición salarial de Oaxaca-Blinder consiste en una técnica para determinar la naturaleza de la diferencia entre salarios. Mediante el análisis de regresión se determinan las brechas en salario, pero con esta técnica lo que se obtiene es cuánto de la brecha total corresponde a la diferencia entre las variables explicativas y cuánto a factores no explicados.

Abreviando la demostración, se toma en cuenta la explicación de Espinoza (2009). Aplicándola a nuestro caso de estudio, las ecuaciones de salario, con las mismas variables de la ecuación anterior para hombres y mujeres, son:

$$\text{Hombre: } W_h = \gamma_h X_h + \theta_h T_h$$

$$\text{Mujer: } W_m = \gamma_m X_m + \theta_m T_m$$

Se obtiene:

$$\text{Hombre: } \bar{W}_h = \hat{\gamma}_h \bar{X}_h$$

$$\text{Mujer: } \bar{W}_m = \hat{\gamma}_m \bar{X}_m$$

La descomposición de Oaxaca-Blinder consiste en la siguiente operación:

$$\begin{aligned} \bar{W}_h - \bar{W}_m &= (\bar{X}_h - \bar{X}_m)\hat{\gamma}_h + (\hat{\gamma}_h - \hat{\gamma}_m)\bar{X}_m \\ &\quad + (\bar{X}_h - \bar{X}_m)(\hat{\gamma}_h - \hat{\gamma}_m) \end{aligned}$$

En el modelo, el primer término de la derecha corresponde a diferencias en las características entre los dos grupos, el segundo término se refiere a diferencias en los rendimientos de esas características, y el tercer elemento es la interacción causada por una diferencia simultánea en las características y retornos. Montenegro (2001) resalta que el segundo componente de la derecha no se explica por el modelo, y se debe a factores no observables (discriminación).

4 Resultados

En el presente estudio, nos enfocamos en la desigualdad salarial por género que ocurre en el mercado laboral, y parcialmente se analiza la desigualdad por grupos étnicos. Podemos subdividir los resultados en tres elementos principales. El primero se refiere a la ecuación de participación de los hombres y mujeres dentro del mercado laboral. Como se indicó, dicha ecuación se utilizó para la corrección del sesgo de selección de Heckman. A continuación, se presentan los resultados concernientes a las estimaciones salariales obtenidas de la ecuación de Mincer, con diferentes variables dicotómicas de interés. Vale la pena resaltar que el coeficiente λ , que es aquel que indica si existen problemas de selección en el mercado laboral, es representativo tanto para hombres como para mujeres; con ello se infiere que, para analizar el mercado laboral ecuatoriano, es necesaria la corrección del sesgo; en su ausencia, se podrían obtener estimaciones sesgadas e inconsistentes. Por último, se presentan los resultados de la descomposición de la brecha salarial usando la metodología Oaxaca-Blinder.

La ecuación de participación se presenta en la Tabla 5, e incluye variables que influyen la decisión de participar en el mercado laboral. Como variable independiente está la probabilidad de participar; como variables independientes, se incluyen: el ser jefe de hogar, presencia de niños menores de 10 años, número de personas en el hogar, y estado civil de la persona.

El ser jefe de hogar influye positivamente en la decisión de participar en el mercado laboral, tanto para hombres como para mujeres. Se podría inferir que la “obligación” de mantenerse a sí mismo y a los suyos, influye en la decisión de trabajar. Para el año 2012, los hombres jefes de hogar tienen un 18.9 % más de posibilidad de participar en el mercado laboral, y las mujeres un 10.9 %.

Además, la presencia de menores de 10 años en el hogar tiene efecto positivo en los hombres hacia la búsqueda de trabajar, y en las mujeres tiene efecto contrario. Esto se debe a que “típicamente” el hombre busca el sustento y la mujer cuida de los niños. Así, para el año 2012, la presencia de niños de hasta 10 años genera en promedio que los

hombres tengan 20.5 % más de posibilidad de insertarse al mercado laboral, y en las mujeres el efecto es de -6.4 %. Por otro lado, el número de personas en el hogar influye negativamente en la participación laboral; así, a mayor número se buscará menos trabajo, aunque en las mujeres no es significativo para el año 2012.

Por último, el estado civil unipersonal influye negativamente en la participación laboral de los hombres; es decir, un hombre que viva casado o en unión tiene una posibilidad de 40.2 % de participar en el mercado laboral, para el año 2012; mientras que esa condición tiene efecto contrario en las mujeres, es decir, las mujeres solteras tienen 66.0 % más de posibilidad de incorporarse al mercado laboral.

Tabla 5. Análisis de Regresión de la Ecuación de Participación

Variable dependiente: participa en el mercado laboral

	2012		2007	
	hombre	mujer	hombre	mujer
jefe de hogar	0.189* (0.046)	0.109* (0.008)	0.425* (0.039)	0.182* (0.053)
menores de 10 años	0.205* (0.023)	-0.064* (0.015)	0.154* (0.023)	-0.047* (0.014)
número de personas	-0.024* (0.010)	-0.004 (0.008)	-0.020* (0.012)	-0.034* (0.008)
estado civil unipersonal	-0.402* (0.045)	0.660* (0.029)	-0.342* (0.051)	0.591* (0.031)
constante	1.088 (0.069)	-0.641* (0.034)	1.118* (0.079)	-0.413* (0.035)

*: significativo al 5 %

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Los resultados de las estimaciones salariales se presentan en la Tabla 6. La ecuación de Mincer utilizada incluye como variable dependiente al logaritmo del salario, y como variables independientes a los años de educación, experiencia y la experiencia al cuadrado. Además, se usa como variables dicotómicas las siguientes: área donde vive, grupo étnico, sector para el cual trabaja, tamaño de la empresa, estudios superiores. Es importante indicar que la endogeneidad es subsanada al usar la corrección del sesgo de selección. Las variables utilizadas fueron elegidas tomando en cuenta la utilidad para comparar los resultados con los obtenidos por las investigaciones que se indicó anteriormente sobre el Ecuador.¹⁰

El estudiar un año adicional genera un incremento positivo en el salario de hombres y mujeres para los dos periodos analizados. Para el año 2012, los hombres tienen en promedio un aumento en su salario de 4.8 % por cada año adicional; las mujeres, un incremento de 6.5 % . Como se

observa, al comparar entre periodos, el retorno disminuye para el año 2012, y se podría pensar que se debe a una mayor educación en general de la sociedad.

Un año más de experiencia tiene efectos positivos en el salario, y la experiencia al cuadrado denota su rendimiento marginal decreciente. El estar ubicados en el área urbana tiene un efecto positivo en el retorno salarial, y ayudaría a explicar la menor densidad en las zonas rurales. Este resultado coincide con el encontrado por Espinoza (2009), aunque en menor medida; las diferencias en estimaciones se podrían explicar porque Espinoza utilizó la Encuesta de Condiciones de Vida 2006 como fuente de datos.

El pertenecer a un grupo minoritario -indígenas, montubios, negros, mulatos o afroecuatorianos- tiene un efecto negativo en el salario, y dicho efecto está más presente en las mujeres. Para el año 2012, los hombres tienen en promedio una reducción en su salario de 3.1 % y las mujeres de 7.0 % por pertenecer a dichos grupos. Este es un indi-

¹⁰En la selección de las variables se usaron otras más, que luego se desecharon por problemas en los coeficientes y significancia de las mismas. Entre ellas, constaban las variables color de piel blanco, región a donde pertenece, entre otras

gador que denota discriminación no solo por género, sino también por grupo étnico. Los resultados concuerdan con los presentados por Córdor (2010) pero en menor magnitud, ya que en dicho trabajo se compara directamente contra indígenas; en nuestro caso se ha construido el grupo minoritario.

Además, se puede apreciar que las personas que trabajan en el sector público tienen un premio salarial tanto para hombres como mujeres. Dicho premio ha aumentado entre periodos y, esto se podría explicar por el fortalecimiento del sector público que ha experimentado el Ecuador durante el último quinquenio. Para el año 2012, los hombres y las mujeres tienen un efecto positivo promedio de 41.8 % y 29.1 % al trabajar en el sector público, respectivamente. Los resultados son similares a los obtenidos por Carrillo (2004) aunque, como se indica, la magnitud de la diferen-

cia a favor del sector público no será la misma, debido a que dicho autor no corrige el sesgo de selección, y también al dinamismo de los últimos años.

Aquellos que trabajan en una micro o pequeña empresa (MIPE)¹¹ tienen un efecto negativo en su salario, con mayor agravio hacia las mujeres. Esta es una cifra interesante ya que, en el Ecuador, la mayoría de empresas son MIPE¹². Adicionalmente, se observa que el realizar estudios de postgrado tiene un impacto positivo en los salarios, con mayor rendimiento para los hombres.

Finalmente, el lambda resulta ser representativo para ambos periodos y entre géneros, indicando que existen problemas de selección; los signos demuestran que, si no se corrige la autoselección, el salario de los hombres estaría sobreestimado y el de las mujeres subestimado.

Tabla 6. Estimaciones Salariales por Género

Variable dependiente: logaritmo del salario

	2012		2007	
	hombre	mujer	hombre	mujer
años de educación	0.048* (0.002)	0.065* (0.002)	0.062* (0.002)	0.089* (0.003)
experiencia	0.013* (0.002)	0.014* (0.002)	0.018* (0.002)	0.022* (0.003)
experiencia 2	-0.000* (0.000)	-0.000* (0.000)	-0.000* (0.000)	-0.000* (0.000)
área urbana	0.135* (0.013)	0.183* (0.021)	0.177* (0.015)	0.138* (0.027)
grupo minoritario	-0.031* (0.016)	-0.070* (0.029)	-0.020* (0.021)	-0.079* (0.037)
sector público	0.418* (0.019)	0.291* (0.025)	0.347* (0.023)	0.125* (0.033)
MIPE	-0.232* (0.015)	-0.305* (0.023)	-0.208* (0.018)	-0.357* (0.029)
postgrado	0.304* (0.052)	0.233* (0.047)	0.410* (0.056)	0.160* (0.074)
lambda	-0.261* (0.049)	0.068* (0.032)	-0.333* (0.063)	0.085* (0.042)
constante	0.145* (0.037)	-0.312* (0.058)	-0.546* (0.043)	-1.149* (0.074)

*: significativo al 5 %

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

¹¹Para clasificar a las empresas en MIPE se utilizó el criterio de número de trabajadores de la Resolución 1260 emitida por la Secretaría de la Comunidad Andina; con ello, una MIPE es una empresa que tiene menos de 50 trabajadores. Es necesario indicar que al trabajar con la ENEMDUR solo se puede usar el criterio de número de trabajadores, ya que no se tiene información de activos ni ventas de la empresa.

¹²Para el año 2012, utilizando la ENEMDUR, cerca del 81 % de empresas es MIPE. Utilizando datos del Censo Nacional Económico 2010, Araque (2012) indica que las MIPE representan cerca del 99 % de empresas en el Ecuador, al usar el criterio de número de trabajadores. Sin embargo, al usar como referencia el aporte de generación de ingresos por ventas, las MIPE reducen su participación a 45 %. Además, se señala que las MIPE contribuyen aproximadamente con el 61 % a la generación del empleo ecuatoriano.

Con el objetivo de determinar las tendencias de las variables para la población en general, en la Tabla 7 se presentan los resultados para los dos periodos que se analizan. Los retornos salariales para el año 2012 se ubican en 5.4 %, y son menores en comparación con el 7% del año 2007. Similarmente, la obtención de un postgrado incrementa el salario en 28.2 % para el año 2012, aunque para el periodo previo era de 30.9 %; la hipótesis de una mayor educación en general de la población explicaría dicho fenómeno. Algo parecido sucede con la experiencia, en la cual, un año adicional otorga un incremento en el salario de 1.4 % para el 2012, inferior al 2.1 % del periodo previo; aquí, dicho efecto se podría explicar por la importancia de manejar nuevos paquetes tecnológicos en el trabajo.

En el mercado laboral, el ser mujer es penalizado con una reducción del 10.1 % en el salario para el año 2012; sin embargo, dicha cifra presenta una disminución al compararla con el 15.1 % del 2007. De igual manera, se observa una reducción en la penalización al salario para las personas que pertenecen a grupos minoritarios, pasando de 6 % en el 2007 a 3.7 % para el año 2012. La explicación de dichas tendencias podría deberse a las políticas de equidad de género y reducción de discriminación hacia grupos minoritarios que se han implementado en los últimos cinco años.

Trabajar para el sector público tiene un efecto positivo de 37.7 % en el salario para el año 2012, y ha aumentado respecto del año 2007 (entonces era de 28.6 %). Como lo mencionamos anteriormente, dicho incremento se explicaría por el fortalecimiento, en materia de recursos destinados hacia los organismos nacionales, municipales, y demás entidades que conforman el sector público.

El vivir en un área urbana genera un incremento del salario de 14.4 %, cifra que se ha reducido al compararla con el periodo anterior. El efecto negativo del 25 % en el salario, por trabajar para una MIPE, se ha mantenido relativamente estable en el tiempo, lo mismo que el efecto del lambda.

Los resultados de la descomposición de la brecha salarial usando el método de Oaxaca-Blinder se indican en la Tabla 8, y las estimaciones son similares a las encontradas en Espinoza (2009). La corrección del sesgo de selección es fundamental para los resultados sobre los diferenciales salariales. Realizando un ejercicio similar al de Fuentes et. al. (2005), y estimando los resultados con corrección y sin ella, se llega a resultados muy diferentes. Con ello, se recalca la importancia de corregir el sesgo de selección.

El primer punto interesante que se observa es la confirmación de la existencia de un diferencial salarial a favor de los hombres para ambos periodos. Comparativamente, entre el periodo 2007 y 2012 se observa una reducción de la brecha salarial. Además, se muestra de forma explícita los componentes de la brecha salarial en los tres elementos antes citados. El primer elemento, correspondiente a diferencias en las características entre los dos grupos, indica en su signo negativo, tanto para 2007 como para 2012-

que las mujeres deberían recibir un salario superior al de los hombres solamente analizando las características. El segundo elemento, referente a las diferencias en los retornos de esas características, al ser positivo y significativo para ambos periodos, muestra que existen factores no observables (discriminación) que determinan el salario entre hombres y mujeres. El tercer elemento, que indica la interacción causada por una diferencia simultánea en las características y retornos, es positivo, y refuerza la brecha salarial por factores no observables.

Tabla 7. Estimaciones Salariales en general

Variable dependiente: logaritmo del salario

	2012	2007
años de educación	0.054* (0.001)	0.070* (0.002)
experiencia	0.014* (0.001)	0.021* (0.002)
experiencia 2	-0.000* (0.000)	-0.000* (0.000)
mujer	-0.101* (0.011)	-0.151* (0.015)
área urbana	0.144* (0.011)	0.164* (0.013)
grupo minoritario	-0.037* (0.014)	-0.060* (0.018)
sector público	0.377* (0.015)	0.286* (0.019)
MIPE	-0.255* (0.012)	-0.251* (0.015)
postgrado	0.282* (0.035)	0.309* (0.045)
lambda	-0.073* (0.019)	-0.075* (0.021)
constante	-0.596* (0.027)	-0.654* (0.036)

*: significativo al 5 %

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Tabla 8. Descomposición salarial de Oaxaca-Blinder

	2012	2007
Diferencia Salarial	0.300*	0.266*
Características	-0.113*	-0.091*
Coefficientes	0.390*	0.348*
Interacción	0.020*	0.009*

*: significativo al 5 %

Fuente: ENEMDUR 2007-2012. Elaboración: Autor.

Con lo encontrado, mediante el uso de la descomposición salarial, se confirma la existencia de discriminación en el mercado laboral ecuatoriano para los años 2007 y 2012. Sin embargo, no existen los elementos suficientes para obtener conclusiones categóricas respecto de dicha discriminación. Este puede ser el punto de partida para futuros estudios.

5 Conclusiones

El enfoque económico al medir la discriminación ha tenido un desarrollo exponencial durante los últimos años. Desde el modelo *taste-based* propuesto por Becker hace ya más de medio siglo, y pionero en dar un tratamiento económico a la medición de la discriminación, se han creado modelos que incorporan diferentes aspectos del mercado. Con la aparición del modelo *statistical-discrimination* de Arrow, Phelps, entre otros, la econometría y el análisis estadístico son las principales herramientas para medir la discriminación.

En el Ecuador, los trabajos sobre brechas salariales son contados, y se han enfocado en la medición de brechas de género, etnia, brechas intersectoriales, brechas intrasectoriales y brechas regionales. Los resultados del presente estudio presentan concordancia con los estudios anteriores en significancia y signo de los coeficientes, pero las magnitudes difieren, ya que se utilizaron diferentes fuentes de datos y metodología. La corrección del sesgo de selección para obtener estimadores insesgados y eficientes es un requisito primordial si se quiere analizar el mercado laboral ecuatoriano. En el estudio, el λ resulta ser representativo para ambos periodos, y los signos demuestran que, si no se corrige la autoselección, el salario de los hombres estaría sobreestimado y el de las mujeres subestimado.

Al analizar la participación laboral, se encontraron patrones significativos por géneros, que influyen en la decisión de participar. Para el año 2012, el ser jefe de hogar influye positivamente tanto para hombres (18.9%) como para mujeres (10.9%). Por otro lado, la presencia de menores de 10 años en el hogar tiene efecto contrario entre géneros: positivo en los hombres (20.5%) y negativo en las mujeres (-6.4%); ocurre algo similar con el estado civil unipersonal, que influye negativamente en la participación laboral de los hombres (-40.2%), y positivamente en las mujeres (66.0%).

El análisis de resultados se enfocó en las brechas salariales entre hombres y mujeres en el mercado ecuatoriano, las mismas que se han reducido de 15.1%, para el año 2007, a 10.1% para el año 2012. La mujer, a pesar de tener más años de educación que los hombres (4.1% más en el 2007, y 7.3% más, en el 2012) recibe menor remuneración promedio que los hombres. Este resultado se agrava si la mujer pertenece a un grupo minoritario, o si trabaja para una MIPE.

Asimismo, las brechas salariales hacia grupos minoritarios han disminuido entre periodos, de 6.0% a 3.7%. En

este caso, en contraste con el anterior, existe una diferencia importante en la formación del capital humano en contra de los grupos minoritarios. Los años de educación promedio de los grupos minoritarios, para el año 2012, son 47.3% menores que los de personas blancas o mestizas. Al analizar por género, el pertenecer a dichos grupos da como resultado, para el año 2012, que los hombres reciben en promedio una reducción en su salario de 3.1% y las mujeres de 7.0%.

En este punto, la hipótesis del autor es que la reducción de brechas responde a las políticas implementadas en los últimos años hacia la mejora en la equidad social, respecto del género como de la pertenencia a grupos minoritarios.. Este sería un tema interesante para abordar en detalle en un estudio posterior.

La descomposición de las diferencias salariales mediante el uso de la metodología de Oaxaca-Blinder demuestra que existe discriminación salarial en el mercado laboral ecuatoriano, debido a que hay factores no observables que determinan la desigualdad. Las brechas salariales, si bien se han reducido entre los periodos analizados, no existen los elementos suficientes para concluir sobre la evolución de la discriminación.

Finalmente, se observa que, a pesar de la mayor participación de la mujer dentro de la sociedad ecuatoriana, todavía existen estigmas de trato no igualitario entre géneros. La política pública es un instrumento ideal para reducir la discriminación salarial. El mejoramiento del capital humano, las campañas de equidad respecto del género y de la pertenencia a grupos minoritarios, y las estructuras organizacionales modernas son tres elementos fundamentales de políticas inclusivas eficientes hacia la reducción de discriminación.

Referencias

- [1] Aigner, Dennis and Glenn Cain. "Statistical Theories of Discrimination in Labor Markets". *Industrial and Labor Relations Review*, 175-187. (1977)
- [2] Araque, Wilson. "Las PyME y su situación actual". Universidad Andina Simón Bolívar, 1-14. (2012)
- [3] Arrow, Kenneth. "Some Mathematical Models of Race in the Labor Market". in *Racial Discrimination in Economic Life*, ed. A.H. Pascal, Lexington: Lexington Books. (1972)
- [4] Arrow, Kenneth. "The Theory of Discrimination". In *Discrimination in Labor Markets*, ed. Orley Ashenfelter and Albert Rees. Princeton, NJ: Princeton University Press, 3-33. (1973)
- [5] Arrow, Kenneth. "What Has Economics to Say about Racial Discrimination?". *Journal of Economic Perspectives*, 91-100. (1998)

- [6] Becker, Gary. "The Economics of Discrimination". 1st edition. Chicago, IL: Chicago University Press. (1957)
- [7] Becker, Gary. "The Economic Way of Looking at Behavior". *Journal of Political Economy*, 385-409. (1993)
- [8] Bertrand, Marianne; Chugh, Dolly; Mullainathan, Sendhil. "Implicit Discrimination". *American Economic Review*, 94-98. (2005)
- [9] Black, Dan. "Discrimination in an Equilibrium Search Model". *Journal of Labor Economy*, 309-334. (1995)
- [10] Cahuc, Pierre and Zylberberg André. "Labor Economics". 1st edition. Cambridge, IL: The MIT Press. (2004)
- [11] Carrillo, Paúl A. y Vasconez, Verónica. "Una Aproximación de la Brecha Salarial entre Entidades Públicas en el Ecuador". *Munich Personal RePEc Archive*, 3-27. (2011)
- [12] Carrillo, Paúl E. "Las Diferencias Salariales entre el Sector Público y Privado en el Ecuador". *Dirección de Estudios del Banco Central del Ecuador*, 2-8. (2004)
- [13] Charles, Kerwin and Guryan, Jonathan. "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination". *Journal of Political Economy*, 773-805. (2008)
- [14] Coate, Stephen, and Loury, Glenn. "Will affirmative action policies eliminate negative stereotypes?". *American Economic Review*, 1220-1240. (1993)
- [15] Córdor, Jorge. "Discriminación salarial en el mercado laboral por etnia". *Facultad Latinoamericana de Ciencias Sociales (FLACSO)*, 25-55. (2010)
- [16] Cooter, Robert. "Market Affirmative Action". *San Diego Law Review*, 133-168. (1994)
- [17] Espinoza, Nereyda. "Estimación de la Brecha Salarial entre Hombres y Mujeres: Un Análisis por Cuantiles para el Ecuador". *Económicas Escuela Superior Politécnica del Litoral (ESPOL)*, 2-20. (2009)
- [18] Fang, Hanming y Moro, Andrea. "Theories of Statistical Discrimination and Affirmative Action: A Survey". *Journal of Economic Literature*, 135-200. (2011)
- [19] Figueroa, María y Miranda, Gustavo. "La rentas laborales a partir de la educación y de la experiencia laboral en el Ecuador periodo 2003-2008, utilizando el modelo de Mincer y las encuesta de empleo y desempleo urbano del INEC (ENEMDU)". *Escuela Politécnica Nacional*, 28-51. (2009)
- [20] Fuentes, Jeanette; Palma, Amalia; Montero, Rodrigo. "Discriminación Salarial por Género en Chile: Una Mirada Global". *Universidad de Chile*, 133-157. (2005)
- [21] Heckman, James. "Sample Selection-Bias as a Specification Error". *Econometrica*, 156-162. (1979)
- [22] Heckman, James. "Detecting Discrimination". *Journal of Economic Perspectives*, 101-16. (1998)
- [23] Jorgensen, Thomas. y Fenger, Anders. "Estimating Discrimination in the Danish Labour Market: a Decomposition of the Gender Wage Gap". *University of Copenhagen*, 3-20. (2009)
- [24] Lang, Kevin. "A Language Theory of Discrimination". *Quarterly Journal of Economics*, 363-382. (1986)
- [25] Lang, Kevin and Lehmann, Jee-Yeon. "Racial discrimination in the Labor Market: Theory and Empirics". *Boston University*, 5-70. (2010)
- [26] Levitt, Steven. "Testing Theories of Discrimination: Evidence from the Weakest Link". *Journal of Law and Economics*, 431-451. (2004)
- [27] List, John. "The Nature and Extent of Discrimination in the Marketplace: Evidence From the Field". *Quarterly Journal of Economics*, 49-89. (2004)
- [28] Loury, Glenn. "Discrimination in the Post-Civil Rights Era: Beyond Market Interaction". *The Journal of Economic Perspectives*, 117-126. (1998)
- [29] Lundberg, Shelly and Startz, Richard. "Private Discrimination and Social Intervention in Competitive Markets". *American Economic Review*, 340-347. (1983)
- [30] McAdams, Richard. "Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination". *Harvard Law Review*, 1005-1084. (1995)
- [31] Mincer, Jacob. "Schooling, Experience and Earnings". *Columbia University Press, New York*. (1974)
- [32] Montenegro, Claudio. "Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach". *Banco Mundial*, 3-35. (2001)
- [33] O'Donnell, Owen; Van Doorslaer, Eddy; Wagstaff Adam; Lindelow, Magnus. "Analyzing Health Equity Using Household Survey Data, A Guide to Techniques and Their Implementation". *Banco Mundial*, 147-157. (2008)
- [34] Phelps, Edmund. "The Statistical Theory of Racism and Sexism". *American Economic Review*, 659-661. (1972)
- [35] Raphael, Steven. "Anatomy of the Anatomy of Racial Inequality". *Journal of Economic Literature*, 1202-1214. (2002)
- [36] Real Academia Española. "Diccionario de la lengua española". Versión en línea. (2013)

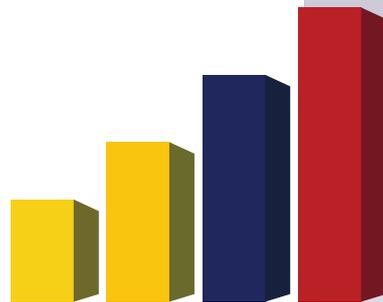
- [37] Sapelli, Claudio. "Ecuaciones de Mincer y las tasas de retorno a la educación en Chile: 1990-1998". Pontificia Universidad Católica de Chile, 2-29. (2003)
- [38] Schwab, Stewart. "Is Statistical Discrimination Efficient?". *American Economic Review*, 228-234. (1986)
- [39] Schwab, Stewart. "Employment Discrimination". Cornell Law School Working Paper No. 5530, 572-592. (1999)
- [40] Sunstein, Cass. "Why Markets Dont Stop Discrimination". *Social Philosophy and Policy*, 22-37. (1991)
- [41] Torres, Victor y Celton, Dora. "Discriminación Salarial en Argentina entre nativos y paraguayos". Universidad de Granada, 263-285. (2009)

Analítica

Factores determinantes de la migración de los
ecuatorianos

Determinants of migration of Ecuadorians

Patricia Cortez y Paúl Medina



www.inec.gob.ec | www.inec.gob.ec/analitika

Factores determinantes de la migración de los ecuatorianos

Determinants of migration of Ecuadorians

Patricia Cortez[†] y Paúl Medina[‡]

[†] Dirección de Estudios Analíticos Estadísticos, Instituto Nacional de Estadística y Censos, Quito, Ecuador

[‡] Departamento de Ciencias Exactas, Escuela Politécnica del Ejército, Quito, Ecuador

[†]patricia_cortez@inec.gob.ec [‡]pmedinavz@gmail.com

Recibido: 20 de noviembre de 2012

Aceptado: 15 de marzo de 2013

Resumen

El presente trabajo surge de la necesidad de explicar el proceso migratorio, sin privilegiar el perfil del migrante, sino poniendo énfasis en las características del hogar de donde provienen. En este trabajo se hace un análisis de las características de los migrantes ecuatorianos a partir de la semejanza existente entre las condiciones sociodemográficas (estructura de la vivienda y servicios en el hogar) de los hogares a los que pertenecían antes de salir del país, considerando el periodo comprendido entre los años 2000 y 2010. Para tal efecto, se realiza un escalamiento óptimo de las variables cualitativas consideradas; posteriormente, se descubren los factores de hogar y vivienda del migrante y, finalmente, se estudian los grupos encontrados. Los datos para el análisis han sido tomados del Censo de Población y Vivienda 2010, elaborado por el Instituto Nacional de Estadística y Censos (INEC).

Palabras clave: proceso migratorio, escalamiento óptimo, condiciones.

Abstract

This work arises from the need to explain the migration process, without favoring the migrant's profile, but with emphasis on household characteristics is from. In this paper we analyze the characteristics of Ecuadorian migrants from the similarity between sociodemographic conditions (structure of housing and services in the home) of the households to which they belonged before leaving the country, considering the period between 2000 and 2010. For this purpose, optimal scaling is performed qualitative variables considered, then the factors are discovered home and migrant housing and ultimately explores the groups found. The data for the analysis are taken from the Census of Population and Housing 2010, prepared by the National Institute of Statistics and Census (INEC).

Keywords: migration process, optimal scaling, sociodemographic conditions.

Código JEL: C45, C61, J13.

1 Introducción

Migrar en busca de mejores condiciones de vida es un fenómeno universal que caracteriza al ser humano. A más de esta condición, el agotamiento sostenido del desarrollo, la economía y la situación política de Ecuador, a partir del año 1999, son algunas de las causas de una migración masiva por parte de los ciudadanos. La superación económi-

ca que experimentaron los primeros migrantes y la mejora en la calidad de vida de sus familiares en el país de origen fueron el detonante: se despertó el interés de una gran cantidad de personas por emprender un viaje para trabajar en el extranjero. Las consecuencias posteriores de éste fenómeno migratorio han sido analizadas en varios trabajos,

véase por ejemplo [6, 9]. Sin embargo; hasta el momento no se ha analizado de manera cualitativa las características sociodemográficas del migrante.

Tradicionalmente, se ha supuesto que las personas que deciden viajar en busca de mejores oportunidades surgen de hogares con condiciones económicas deficientes. En el trabajo desarrollado por J. E. Torres [10] se evidencia una clara relación entre la pobreza y las condiciones habitacionales del hogar. Se menciona que la vivienda forma parte del ingreso del individuo o de la familia. Cuando hay desempleo o subempleo, el ingreso es bajo; en consecuencia, no se contará con los recursos necesarios para cubrir las necesidades básicas. Fundamentalmente, se verán afectadas las condiciones de la vivienda, ya que ésta representa, en la mayoría de los casos, el bien más costoso.

Desde esta perspectiva, se evaluará el fenómeno migratorio en el Ecuador, considerando el contexto del ambiente familiar del migrante. Se busca, así, encontrar similitudes en las características de los hogares de origen para, en base a los resultados, inferir una tendencia de comportamiento.

A continuación, se describe el orden de presentación de las secciones que conforman la investigación. La sección 2 inicia con una descripción de las técnicas de análisis multivariante utilizadas en el estudio. En la sección 3, se describe la metodología utilizada para la obtención de los resultados. En la sección 4, se presentan los resultados, es decir, los grupos de migrantes de acuerdo a las características del hogar de origen, producto de la aplicación de los modelos descritos en la sección 2. Finalmente, en la sección 5, se exponen las principales conclusiones de la investigación realizada.

2 Marco Teórico

Generalmente, el concepto de migración se asocia con el desplazamiento temporal o definitivo de un lugar a otro. En los seres humanos, la migración se define como el proceso de movilidad regional, dentro de una misma sociedad o entre distintas sociedades. El acto de migrar -sea de manera voluntaria o forzada- se relaciona, en la mayoría de las ocasiones con razones económicas o políticas. Por otro lado, explicar parcial o completamente este fenómeno implica reunir elementos de los diversos sistemas que conforman a la sociedad; de allí que sea necesario contar con herramientas estadísticas adecuadas para abordarlo.

En vista de lo expuesto, y dadas las características de las variables involucradas en el análisis, se hace necesario recurrir a la técnica de análisis multivariante denominada análisis factorial; ello, previo a un escalamiento óptimo de las variables, que serán presentadas en la siguiente sección; finalmente, se concluirá con un análisis de conglomerados de K medias.

2.1 Escalamiento óptimo

Ante la necesidad de trabajar con variables cuantitativas nació un conjunto de técnicas basadas en lo que se denomina escalamiento óptimo; esto consiste en asignar un valor escalar a las categorías de variables cualitativas, es decir, este proceso convierte variables cualitativas o categóricas en variables cuantitativas o de escala. Los datos de partida son, entonces, variables cualitativas que describen a los individuos mediante un número limitado de categorías que no tienen un punto inicial de referencia; por otro lado, se desconoce la medida de distancia entre las categorías que, además, pueden ser desordenadas.

Los valores óptimos de escala son relacionados con las categorías en un proceso basado en un criterio de optimización, dependiendo del procedimiento de análisis de datos para el cual serán utilizadas las variables. Los valores de escala finales tienen propiedades métricas a diferencia de las variables categóricas.

La cuantificación óptima se obtiene, a través de un método iterativo realizado en dos pasos, que tiene por nombre mínimos cuadrados alternantes; con él, utilizando las asignaciones previas para estimar las siguientes, se llega a la solución. En el primer paso, se calcula una base óptima para los valores dados de la transformación; en el segundo, los nuevos valores para las transformaciones óptimas de la base calculada en el primer paso. Las transformaciones son escogidas de tal manera que minimicen la función de pérdida. Cuando se tienen varias variables de interés, el proceso consiste en cuantificar las categorías de las variables de tal manera que se maximicen las correlaciones entre todas ellas. Para más detalles véase [5].

2.2 Análisis factorial

El principal objetivo de este análisis es simplificar la información que brinda una matriz de correlaciones con el fin de hacerla de más fácil interpretación. Se aplica en casos donde es necesario recoger medidas indirectas que estén relacionadas con los conceptos que interesan. Las variables que interesan reciben el nombre de *variables latentes*; la metodología que las relaciona con variables observadas recibe el nombre de *análisis factorial*.

El Análisis factorial busca esencialmente nuevas variables o *factores*, tratando de explicar la estructura de las covarianzas entre las variables; para ello, es necesario que las variables originales no estén incorreladas¹, porque si lo estuvieran, no habría nada que explicar de las variables.

Se considera un conjunto de p variables observadas x_1, x_2, \dots, x_p que se asume relacionadas con un número dado de variables latentes f_1, f_2, \dots, f_k , donde $k \ll p$, mediante una relación del tipo

$$X = \Lambda F + U, \quad (1)$$

¹La correlación nula se da cuando no hay dependencia de ningún tipo entre las variables. En este caso se dice que las variables son incorreladas.

donde,

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pk} \end{pmatrix},$$

$$F = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{pmatrix} \text{ y } U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}.$$

Los λ_{ij} son los pesos factoriales que muestran como cada x_i depende de factores comunes; se usan para interpretar los factores. Se asume que los términos residuales u_1, u_2, \dots, u_p , denominados variables específicas, están incorrelados entre sí y con los factores f_1, f_2, \dots, f_k .

Dado que los factores no son observables. Se puede fijar arbitrariamente su media en 0 y su varianza en 1, esto es, se consideran variables estandarizadas que están incorreladas entre sí; de modo que los pesos factoriales resultan ser las correlaciones entre las variables y los factores.

Cuando se consigue una estructura simple, las variables observadas se encuentran en grupos mutuamente excluyentes, de modo que los pesos son altos en unos pocos factores y bajos en el resto.

Hay dos posibles tipos de rotaciones: *ortogonales* y *oblicuas*. La ventaja principal de las rotaciones ortogonales es su simplicidad; ya que los pesos representan las correlaciones entre los factores y las variables, sin embargo esto no se cumple en el caso de las rotaciones oblicuas. Entre las rotaciones ortogonales se encuentran dos tipos principales: *Varimax* y *Cuartimax*. Por otra parte, entre las rotaciones oblicuas, la más empleada es la denominada *Oblimín*.

Para más detalles al respecto véase [11].

2.3 Análisis de conglomerados

El análisis de conglomerados es una técnica multivariante que permite agrupar los casos de un conjunto de datos en función de la similaridad existente entre ellos. Existen dos métodos en esta técnica: el análisis de conglomerados jerárquico y el análisis de conglomerados de K medias. El método jerárquico es apto para determinar el número óptimo de conglomerados existente en los datos; mientras que, el método de K medias permite procesar un número ilimitado de casos pero requiere, que se conozca de antemano el número de conglomerados que se desea obtener.

Ambos métodos son de tipo *aglomerativo*, es decir, partiendo del análisis de los casos individuales, se agrupa los casos hasta llegar a la formación de conglomerados homogéneos.

2.3.1 Análisis de conglomerados jerárquico

El criterio básico para establecer una agrupación es la distancia. Los objetos que estén cerca uno del otro pertene-

cen al mismo conglomerado, y viceversa. Para un conjunto de datos dado, los conglomerados que se construyen dependen de la especificación de los siguientes criterios:

- El método de conglomerado define las reglas para la formación del mismo. Por ejemplo, cuando se calcula la distancia entre dos conglomerados, se puede usar el par de objetos más cercado entre conglomerados o el par de objeto más alejados, o un enlace entre los dos.
- La fórmula para el cálculo de la distancia. Por ejemplo, la medida de la distancia Euclídea.

La versatilidad del análisis de conglomerados jerárquico radica en la posibilidad de utilizar distintos tipos de medidas para estimar la distancia existente entre los casos; sin embargo, no existe ninguna combinación óptima de estas posibilidades para obtener la solución. En general, es conveniente experimentar con distintas soluciones para elegir la más apropiada.

2.3.2 Análisis de conglomerados de K medias

Este análisis empieza con la construcción de centros de conglomerados iniciales. El investigador puede escoger estos centros o tener un procedimiento de selección de k observaciones bien definidas para los centros de los conglomerados. Luego de la obtención de los centros de los conglomerados, el procedimiento consta de los siguientes pasos:

- Asignar casos a los conglomerados basándose en la distancia de los centros de los conglomerados.
- Actualizar las posiciones de los centros de los conglomerados basándose en los valores medios de los casos en cada conglomerado.

Estos pasos se repiten hasta que cualquier reasignación de los casos haga que los conglomerados sean internamente más variables o externamente similares.

Una visión más detallada del análisis de conglomerados jerárquico y de K medias se puede obtener en [12, 13], respectivamente.

3 Metodología

A continuación, se procede a describir la metodología utilizada para la obtención de los resultados buscados. En primer lugar, cabe mencionar que la fuente de datos está constituida por el Censo de Población y Vivienda del año 2010, específicamente, la sección titulada Remesas y Migración.

En [2] se sintetizan las principales variables reconocidas en la literatura sobre migración internacional. Se muestra que, desde un punto de vista económico, el proceso está motivado por la demanda de mano de obra por parte de

los países de destino, o bien, por decisiones individuales que tienen como objetivo la mejora de una calidad de vida. Del mismo modo, en, [4] se muestra el grado de influencia que tienen el sexo y la escolaridad del jefe del hogar en los desplazamientos migratorios, así como las características de los hogares que favorecen la movilidad de la población.

Tales argumentos, así como la perspectiva de los autores, sirvieron de plataforma para la selección de las variables involucradas. En la Tabla 1 se resumen las variables que se escogieron para el desarrollo de la investigación. Cabe recalcar que dichas variables se refieren al hogar del cual salió la persona migrante.

De vivienda	De hogar	De jefe de hogar	De migración
<ul style="list-style-type: none"> ▪ Tipo de vivienda ▪ Vía de acceso principal ▪ Material predominante del techo ▪ Material predominante de las paredes ▪ Material predominante del piso ▪ Estado del techo ▪ Estado de las paredes ▪ Estado del piso ▪ Procedencia principal del agua ▪ Recepción del agua ▪ Servicio higiénico ▪ Servicio de luz ▪ Eliminación de basura 	<ul style="list-style-type: none"> ▪ Exclusividad de servicio higiénico ▪ Exclusividad de instalaciones para bañarse ▪ Disponibilidad de teléfono convencional ▪ Disponibilidad de servicio de internet ▪ Disponibilidad de computadora ▪ Disponibilidad de televisión por cable ▪ Exclusividad de cuarto para cocinar ▪ Combustible para cocinar 	<ul style="list-style-type: none"> ▪ Sexo ▪ Nivel de instrucción 	<ul style="list-style-type: none"> ▪ Provincia ▪ Área ▪ Sexo ▪ Edad al salir del país ▪ Año de salida ▪ Actual país de residencia ▪ Principal motivo de viaje

Tabla 1. Variables utilizadas para el análisis. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Como se observa, las variables de vivienda y hogar suman 21 en total y, además, son todas cualitativas. Por tanto, se hace necesario condensar la información en un número más pequeño de variables, mediante el análisis factorial; para ello, se necesita realizar antes un escalamiento óptimo de las mismas. En este punto cabe señalar que se prefirió el análisis factorial en lugar del análisis de componentes principales (ACP), en razón de lo que se analiza y se persigue con cada técnica; en el análisis factorial, se estudia la estructura de las correlaciones entre las **variables**, se buscan factores hipotéticos que **expliquen** las variables originales y se representa las **correlaciones** entre variables y, entre variables y factores; mientras que en el ACP se estudia la información de los **individuos**, se describen los valores de los individuos mediante un pequeño número de variables que sean combinación de las originales, y se representan **individuos**. Por tales razones, la técnica que responde a los intereses de la investigación es el análisis factorial.

No se realizó el mismo procedimiento para las variables de jefe de hogar y migración. En el primer caso, apenas se cuenta con dos variables, por lo que el análisis factorial

resultaría innecesario. En el segundo, puesto que se quiere examinar el fenómeno migratorio, se considera que es importante rescatar todas las variables, sin abreviarlas en factores.

Una vez obtenidos los factores de vivienda y hogar, es necesario definir el número óptimo de conglomerados. Sin embargo, se presenta un inconveniente. Se cuenta con 280.374 registros, lo que hace imposible aplicar, directamente, el análisis de conglomerados jerárquico, pues esta técnica está limitada a un número pequeño de datos. Por otra parte, el análisis de conglomerados de K medias requiere que se conozca de antemano el número de conglomerados que se va a formar.

Por tanto, es necesario obtener un número reducido de observaciones que sea representativo de la población. Se hace uso de la ecuación (2) para determinar el tamaño apropiado de la muestra con un error muestral del 3%.

$$n = \frac{z^2 N p q}{e^2 (N - 1) + z^2 p q}, \quad (2)$$

donde N es el tamaño de la población, z se obtiene de la

distribución normal estándar, e es el error muestral deseado, en tanto por uno. En cuanto a p y q son datos generalmente desconocidos; se suele suponer que $p = q = 0,5$.

Es así que se obtuvo un tamaño muestral de 1302 registros que representa aproximadamente el 0,5 %, de la base original, para de ella determinar el número óptimo de conglomerados a partir del análisis de conglomerados jerárquico; dicho número servirá para aplicar el análisis de conglomerados de K medias a todos los registros.

4 Resultados

En primer lugar, se realizó un análisis factorial bajo la metodología de componentes principales con todas las variables de vivienda y de hogar, por separado; el objeto es descubrir la estructura que comparten entre sí las variables. Los factores encontrados se caracterizan por abreviar las variables que tienen una correlación alta entre sí y escasa correlación con el resto que forman otro factor. Para su aplicación se ha procedido a la transformación de las variables cualitativas en cuantitativas. Este procedimiento, además, estandariza las variables involucradas, lo que facilita su comparabilidad, (véase [3]).

El elevado tamaño muestral con el que se cuenta posibilitó la obtención de estimaciones confiables en los coeficientes de correlación. A partir de la matriz de correlacio-

nes se obtuvo, en el caso de las variables de vivienda, un *determinante* de 0,010, una medida de adecuación muestral de *Kaiser-Meyer Olkin* de 0,890 y una prueba de esfericidad de Bartlett de 1281426,953 con 78 grados de libertad y significancia estadística de 0,000. Según lo expuesto en [11], estos estadísticos indican que el análisis factorial es factible, ya que existe la suficiente varianza común entre las variables observadas para permitir su agrupación en combinaciones lineales de variables correlacionadas. Del mismo modo, para las variables de hogar, se obtuvo un *determinante* de 0,196, una medida de adecuación muestral de *Kaiser-Meyer Olkin* de 0,665 y una prueba de esfericidad de Bartlett de 457586,174 con 28 grados de libertad y significancia estadística de 0,000.

Tanto para las variables de hogar como para las variables de vivienda, la matriz factorial se rotó mediante el procedimiento de rotación ortogonal *Varimax*; se busca con ello mayor simplicidad en la solución, pues se considera máxima la varianza de los coeficientes factoriales cuadrados en cada componente principal de modo que se obtienen factores no correlacionados entre sí, (véase [11]).

Las Tablas 2 y 3 resumen la composición de los factores obtenidos tras la aplicación del criterio de los autores de maximizar la varianza explicada por el modelo. Se ha intentado que el nombre dado a cada factor refleje el significado de las variables que en él pesan más y que esquematice convenientemente su contenido común.

	Factores de vivienda				
	Estado de la vivienda	Servicios de higiene	Material de la vivienda	Servicio de agua	Servicio de luz
Estado de las paredes	,840				
Estado del piso	,821				
Estado del techo	,818				
Vía de acceso principal		,796			
Eliminación de basura		,786			
Servicio higiénico		,658			
Tipo de vivienda			,798		
Material predominante del techo			,684		
Material predominante de las paredes			,639		
Material predominante del piso			,466		
Procedencia principal del agua				,818	
Recepción del agua				,792	
Servicio de luz					,972

Tabla 2. Matriz de factores rotados para las variables de vivienda. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

	Factores de hogar			
	Servicios comunes	Exclusividad de instalaciones	Exclusividad de cocina	Combustible
Disponibilidad de internet	0,832			
Disponibilidad de computadora	0,804			
Disponibilidad de televisión por cable	0,614			
Disponibilidad de teléfono convencional	0,547			
Exclusividad de servicio higiénico		0,924		
Exclusividad de instalaciones para bañarse		0,910		
Exclusividad de cuarto para cocinar			0,971	
Combustible para cocinar				0,978

Tabla 3. Matriz de factores rotados para las variables de hogar. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

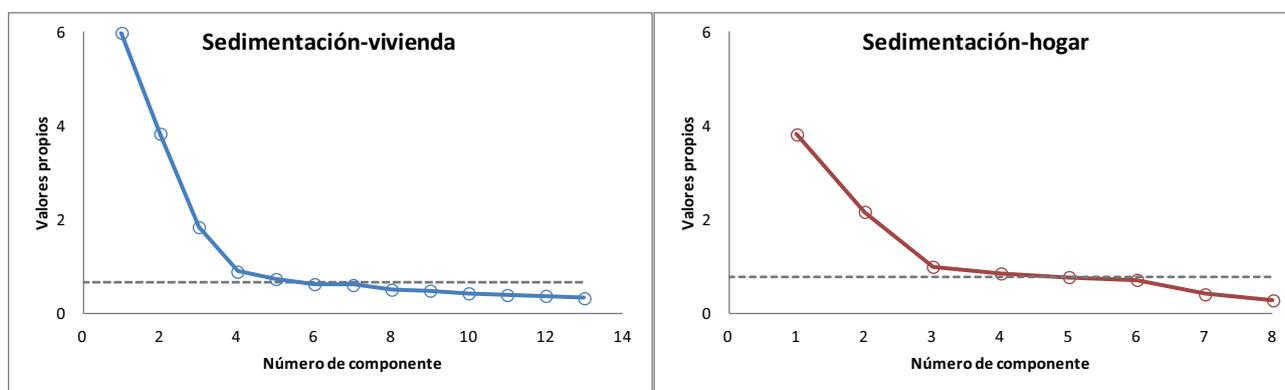


Figura 1. Gráficos de sedimentación para las variables de vivienda y de hogar. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

La varianza total explicada por la combinación de las 13 variables empíricas de vivienda en 5 factores es de 78,018 %; mientras que para las 8 variables de hogar, se tiene un 78,251 % con 4 factores. Es habitual que la proporción de varianza total explicada sea mayor en los primeros factores y menor en los últimos. La Figura 1 muestra los gráficos de sedimentación de los valores propios de la matriz de correlación, los mismos que indican la varianza total explicada, en cada caso.

La solución factorial deja como resultado adicional las comunalidades finales de las variables, es decir, la proporción de varianza de cada variable que es explicada por los factores encontrados. En la Tabla 4 se observa que las variables, para el caso de vivienda, cuya variabilidad queda menos explicada son material predominante del techo y material predominante de las paredes. Mientras que, para el caso de hogar, la Tabla 5 refleja que las variables disponibilidad de teléfono convencional y disponibilidad de televisión por cable tienen la variabilidad más baja.

Concluido el análisis factorial, se procede a realizar el análisis de conglomerados jerárquico, para determinar el número óptimo de agrupaciones. Debido a la gran cantidad de registros existente, se reducen los registros a una

muestra que represente aproximadamente el 0,5 % de los datos originales, tomados de manera aleatoria de acuerdo a lo señalado anteriormente en la determinación del tamaño de la muestra. Es decir, este valor tiene un error muestral de aproximadamente el 0,03 %.

Variables de vivienda	Comunalidad
Tipo de vivienda	0,690
Vía de acceso principal	0,685
Material predominante del techo	0,585
Material predominante de las paredes	0,595
Material predominante del piso	0,601
Estado del techo	0,713
Estado de las paredes	0,754
Estado del piso	0,736
Procedencia principal del agua	0,783
Recepción del agua	0,763
Servicio higiénico	0,665
Servicio de luz	0,981
Eliminación de basura	0,709

Tabla 4. Comunalidades finales para las variables de vivienda. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

VARIABLES DE HOGAR	COMUNALIDAD
Exclusividad de servicio higiénico	0,863
Exclusividad de instalaciones para bañarse	0,857
Disponibilidad de teléfono convencional	0,517
Disponibilidad de servicio de internet	0,697
Disponibilidad de computadora	0,663
Disponibilidad de televisión por cable	0,592
Exclusividad de cuarto para cocinar	0,962
Combustible para cocinar	0,975

Tabla 5. Comunalidades finales para las variables de hogar. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Cabe aclarar que se realizó previamente una estandarización de las variables, puesto que tienen medidas distintas. De este modo, la muestra usada cuenta con 1302 registros.

Para determinar el número apropiado de conglomerados, se hace uso del *dendrograma*, que es el resultado de la aplicación del análisis de conglomerados jerárquico. Él muestra que una buena opción para aplicar el análisis de conglomerados de K medias es tomar 6 conglomerados (véase Figura 2). A continuación, se exhiben los resultados de la aplicación de este análisis.

La Tabla 6 contiene los centros iniciales, es decir, los valores que toman las variables utilizadas en el análisis y que han sido elegidos como centros respectivos de los conglomerados solicitados. Al estar las variables estandarizadas, su interpretación se hace en unidades de desviación típica por encima (signo positivo) o por debajo (signo negativo) de la media.

Una vez seleccionados los centros de los conglomerados, cada caso es asignado al conglomerado de cuyo centro se encuentra más próximo y comienza el proceso de ubicación iterativa de los centros. En cada iteración, se reasignan los casos por su distancia respecto del nuevo centro. La Tabla 7 resume el historial de iteraciones, 9 para este caso, indicando el cambio experimentado por los centros en cada iteración. Puede observarse que, a medida que avanzan

las iteraciones, el desplazamiento de los centros se hace cada vez más pequeño, hasta llegar a la última iteración en la que no existe cambio alguno.

Por otra parte, la Tabla 8 ofrece los centros de los conglomerados finales, es decir, los centros de los conglomerados tras el proceso de actualización iterativa. Al comparar los centros finales con los centros iniciales se puede apreciar con claridad un cambio en los centros de cada conglomerado. Esta información resulta de gran utilidad, pues resume los valores centrales de cada conglomerado en las variables de interés.

En cuanto al número de casos en cada conglomerado se tiene que, el más numeroso es el conglomerado 3, seguido del 4, mientras que los conglomerados con menos casos están dados por el 2 y el 6 (véase Tabla 9).

Finalmente, puesto que los grupos se han formado deliberadamente tomando en cuenta la distancia entre ellos en un espacio multidimensional (que no se evidencia la condición de aleatoriedad en las observaciones en los diferentes grupos), los resultados del análisis son puramente descriptivos.

En este sentido, para verificar la hipótesis de que la varianza entre los grupos es igual, no se puede usar el nivel de significancia que se presenta en la Tabla 10, y que muestra el análisis de dispersión de los resultados de clasificación, para verificar la hipótesis de que la varianza entre los grupos es igual. Sin embargo, la columna F de la Tabla ANOVA hace posible extraer conclusiones generales sobre el comportamiento de las variables en la formación de los *clusters*. Esto muestra que el sexo del migrante y el actual país de residencia son las variables que menos influyen en la formación de los conglomerados; por el contrario, la exclusividad de las instalaciones en el hogar y el servicio de luz son las variables que tienen mayor influencia en los conglomerados.

Una vez establecidos los conglomerados, se procede a realizar un análisis de cada uno de ellos; el objetivo es determinar si existe una relación entre las condiciones estructurales del hogar de donde salió el migrante, y el posible motivo por el cual abandonó el país.

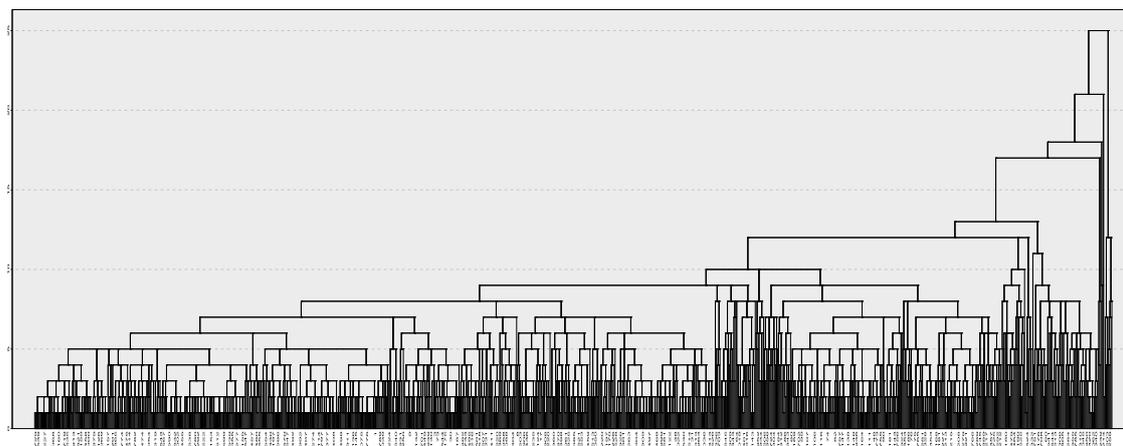


Figura 2. Dendrograma del análisis de conglomerados jerárquico. Fuente: Herramienta estadística SPSS a partir del Censo de Población y Vivienda 2010.

Variables	Conglomerado					
	1	2	3	4	5	6
Estado de la vivienda	,52891	-,89937	-,66548	-,91800	-,73206	,74931
Servicios de higiene	-,18850	-1,87074	-2,77255	-,50658	-,41843	1,51481
Material de la vivienda	7,14583	-1,90457	7,91691	-,16704	-,96193	-1,37566
Servicio de agua	-,99086	4,07722	3,53577	-,23643	-,12871	1,50285
Servicio de luz	6,57786	7,11204	-1,38615	,10622	,09093	-,26727
Servicios comunes	,40037	,68901	-,48145	-1,84602	-2,96496	1,03810
Exclusividad de instalaciones	,73668	4,02633	1,78969	-,03516	3,87069	,22754
Exclusividad de cocina	2,82058	-,70733	-,78359	,21942	2,66470	-,26756
Combustible	4,07902	-,34949	-,32503	-3,22542	4,32421	-,14717
Sexo del jefe de hogar	-,78016	-,78016	1,28179	1,28179	1,28179	1,28179
Nivel de instrucción del jefe de hogar	-1,90184	-,19981	-,62532	1,92772	1,50222	-,19981
Provincia	1,13867	,01420	-1,59219	,97804	-,30708	12,70470
Área	1,59563	-,62671	-,62671	1,59563	-,62671	1,59563
Sexo del migrante	-,92627	-,92627	1,07960	1,07960	1,07960	-,92627
Edad al salir del país	-,22292	-,92933	,40500	-,77235	-1,00782	3,93705
Año de salida	,13182	,46861	-1,21530	-,87852	-1,21530	1,81573
Actual país de residencia	,71154	,71154	-1,18765	6,35153	-1,18765	-1,18765
Principal motivo de viaje	-,67026	-,67026	-,67026	,37053	2,45211	2,45211

Tabla 6. Centros iniciales de los conglomerados. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Iteración	Cambio en los centros de los conglomerados					
	1	2	3	4	5	6
1	8,886	6,807	10,477	7,519	8,015	5,902
2	5,766	,234	1,193	,836	3,809	,023
3	,069	,001	2,866	2,965	,033	9,20E-05
4	,001	1,99E-07	,012	,044	,005	3,82E-07
5	9,79E-08	3,37E-11	6,59E-08	7,92E-07	2,79E-07	1,66E-09
6	9,47E-12	4,92E-15	3,52E-13	1,42E-11	1,53E-11	7,52E-12
7	9,17E-16	0,00E+00	0,00E+00	7,24E-16	7,48E-16	3,51E-14
8	,000	,000	,000	,000	,000	2,56E-16
9	,000	,000	,000	,000	,000	,000

Tabla 7. Historial de iteraciones. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Factores determinantes de la migración de los ecuatorianos

Variables	Conglomerado					
	1	2	3	4	5	6
Estado de la vivienda	,54336	,29958	-,15851	,23775	,48622	,14270
Servicios de higiene	1,24457	,61028	-,37788	1,02246	-,20275	,69405
Material de la vivienda	,42491	,76987	-,12168	,16411	,24523	,46630
Servicio de agua	,49811	,79778	-,17266	,40352	-,03153	,94048
Servicio de luz	-,37956	6,52965	-,06797	-,33452	-,16834	,60668
Servicios comunes	,42774	,69186	-,25927	,67919	,08619	,54467
Exclusividad de instalaciones	,22377	,42244	-,31781	-,08436	3,28188	,12281
Exclusividad de cocina	,28187	,63519	-,04164	-,01086	,09179	,25796
Combustible	4,33458	1,02493	-,17162	-,20918	-,40609	,12949
Sexo del jefe de hogar	-,16950	-,07638	-,01031	,06434	,03090	-,14926
Nivel de instrucción del jefe de hogar	-,67207	-,53160	,29137	-,69671	-,29640	-,46755
Provincia	-,27054	-,04716	,13225	-,47537	,09765	12,70470
Área	,96327	,52905	-,48000	1,40521	-,16374	1,59563
Sexo del migrante	-,09189	-,01500	,06658	-,19929	-,01155	,10660
Edad al salir del país	-,12324	-,09177	,05246	-,12832	-,04080	-,07385
Año de salida	,02704	,03736	-,00321	,03200	-,09139	-,11606
Actual país de residencia	-,11287	,07770	,09208	-,35244	,18512	,00205
Principal motivo de viaje	-,18523	,01687	,10544	-,29792	-,06585	,31616

Tabla 8. Centros finales de los conglomerados. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Conglomerado	Número	Porcentaje
1	10508,000	3,75 %
2	5930,000	2,12 %
3	188618,000	67,27 %
4	56814,000	20,26 %
5	18236,000	6,50 %
6	268,000	0,10 %

Tabla 9. Número de casos en cada conglomerado. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

Variables	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
Estado de la vivienda	3180,373	5	,943	280368	3371,539	,000
Servicios de higiene	21138,554	5	,623	280368	33928,134	,000
Material de la vivienda	2177,923	5	,961	280368	2265,891	,000
Servicio de agua	4302,114	5	,923	280368	4659,522	,000
Servicio de luz	52438,315	5	,065	280368	808628,800	,000
Servicios comunes	8772,738	5	,844	280368	10399,569	,000
Exclusividad de instalaciones	43491,737	5	,224	280368	193814,266	,000
Exclusividad de cocina	746,536	5	,987	280368	756,596	,000
Combustible	42942,495	5	,234	280368	183362,851	,000
Sexo del jefe de hogar	12070,320	5	,785	280368	15380,920	,000
Nivel de instrucción del jefe de hogar	33644,879	5	,400	280368	84111,144	,000
Provincia	637,620	5	,989	280368	644,942	,000
Área	339,170	5	,994	280349	341,228	,000
Sexo del migrante	39,707	5	,999	251131	39,737	,000
Edad al salir del país	1792,409	5	,966	266416	1854,767	,000
Año de salida	1521,595	5	,973	280368	1564,007	,000
Actual país de residencia	123,030	5	,998	280368	123,298	,000
Principal motivo de viaje	10140,898	5	,816	275770	12425,241	,000

Tabla 10. Análisis de varianza ANOVA. Fuente: Elaboración propia a partir del Censo de Población y Vivienda 2010.

4.1 Conglomerado 1

El grupo representativo de migrantes pertenecientes al conglomerado 1 está constituido por mujeres que salieron del área urbana de la provincia de Guayas, en una edad comprendida entre 11 y 20 años. Emigraron del Ecuador en el año 2002 con destino hacia España. Provenían de un hogar donde el jefe era un hombre, que alcanzó la educación primaria. El hogar dispone de servicios básicos, el estado y el material de la vivienda son buenos, y hay exclusividad de instalaciones. Sin embargo, el hogar no dispone de bienes y servicios comunes apropiados; y el principal posible motivo de viaje de este grupo es la unión familiar.

4.2 Conglomerado 2

El grupo representativo de migrantes pertenecientes al conglomerado 2 está constituido por hombres que salieron del área urbana de la provincia de Pichincha, en una edad comprendida entre 21 y 30 años. Salieron del Ecuador en el año 2002 con destino a España. Provenían de un hogar donde el jefe era un hombre que alcanzó la educación primaria. El hogar dispone de servicios básicos, el estado de la vivienda es regular, el material es bueno. Sin embargo, aunque el hogar dispone de exclusividad de instalaciones; estas son deficientes, además no dispone de bienes y servicios comunes apropiados; el principal posible motivo de viaje de este grupo es el trabajo.

4.3 Conglomerado 3

El grupo representativo de migrantes pertenecientes al conglomerado 3 está constituido por hombres que salieron del área urbana de la provincia de Pichincha, en una edad comprendida entre 21 y 30 años. Salieron del Ecuador en el año 2002 con destino hacia España. Provenían de un hogar donde el jefe era un hombre que alcanzó la educación superior. El hogar dispone de servicios básicos, el estado y el material de la vivienda son buenos; con exclusividad de instalaciones y disponibilidad de bienes y servicios comunes apropiados. El principal posible motivo de viaje de este grupo es el trabajo.

4.4 Conglomerado 4

El grupo representativo de migrantes pertenecientes al conglomerado 4 está constituido por hombres que salieron del área rural de la provincia de Azuay, en una edad comprendida entre 21 y 30 años. Emigraron del Ecuador en el año 2002 con destino hacia Estados Unidos. Provenían de un hogar donde el jefe era un hombre que alcanzó la educación primaria. El hogar dispone de servicios básicos deficientes, el estado de la vivienda es regular, el material es deficiente y con exclusividad de instalaciones. Sin embargo, el hogar no dispone de bienes y servicios comu-

nes apropiados; el principal posible motivo de viaje de este grupo es el trabajo.

4.5 Conglomerado 5

El grupo representativo de migrantes pertenecientes al conglomerado 5 está constituido por hombres que salieron del área rural de las Zonas no Delimitadas, en una edad comprendida entre 21 y 30 años. Emigraron del Ecuador en el año 2002 con destino hacia España. Provenían de un hogar donde el jefe era un hombre que alcanzó la educación primaria. El hogar dispone de servicios básicos, el estado de la vivienda es regular, el material es deficiente y con exclusividad de instalaciones. Sin embargo, el hogar no dispone de bienes y servicios comunes apropiados; el principal posible motivo de viaje de este grupo es el trabajo.

4.6 Conglomerado 6

El grupo representativo de migrantes pertenecientes al conglomerado 6 está constituido por hombres que salieron del área urbana de la provincia de Guayas, en una edad comprendida entre 21 y 30 años. Salieron del Ecuador en el año 2001 con destino hacia España. Provenían de un hogar donde el jefe era un hombre que alcanzó la educación primaria. El hogar dispone de servicios básicos, el estado y el material de la vivienda son buenos y con exclusividad de instalaciones. Sin embargo, el hogar no dispone de bienes y servicios comunes apropiados; el principal posible motivo de viaje de este grupo es el trabajo.

5 Conclusiones

Pueden extraerse varias conclusiones de esta investigación. Entre las principales se destaca que:

- Las personas que salieron del país por unión familiar, en su mayoría eran mujeres.
- Las provincias de origen de la mayoría de los migrantes son Pichincha, Guayas y Azuay; ellas albergan a las principales ciudades del país.
- Se refleja un conglomerado, el de menor número de casos, con migrantes provenientes de las Zonas no Delimitadas; en este caso, se puede considerar a esta agrupación como un hecho aislado dentro del fenómeno migratorio.
- En cinco de los seis conglomerados que se formaron, el jefe de hogar de origen de los migrantes estaba constituido por hombres que apenas alcanzaron la educación primaria. Esto confirma lo planteado por otros estudios; que la migración y la escolaridad del jefe de hogar tienen una elevada correspondencia.

- Los principales destinos de los migrantes fueron España y Estados Unidos. Además, el fenómeno migratorio se aceleró entre los años 2001 y 2002.
- Finalmente, se comprueba que las condiciones estructurales del hogar se relacionan directamente con el motivo de la migración: quienes salieron por reunificación familiar, disponían de mejores características distributivas en sus hogares que aquellos que lo hicieron por trabajo, es decir, por buscar mejores condiciones económicas.

Referencias

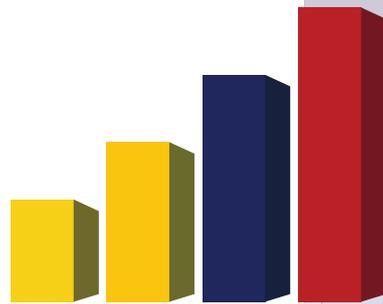
- [1] Álvarez, R. (1995) "Estadística Multivariante y no paramétrica con SPSS". España: Díaz de Santos.
- [2] Brettell, C. y Hollifield, J. (2000) "Migration Theory. Talking a cross disciplines". New York, Routledge, 1-26.
- [3] Cea DANcona, M. (2001) "Metodología Cuantitativa. Estrategias y técnicas de investigación social". Madrid: Síntesis.
- [4] Chávez, A. y Serrano, O. (2003) "La migración reciente en hogares de la región centro de México". Papeles de Población, Vol. 36, 79-108. ISSN, 1405-7425.
- [5] Correa, G. (2008). Contribuciones al análisis multivariante no lineal. Tesis de Doctorado, Departamento de Estadística de la Universidad de Salamanca, Salamanca, España.
- [6] Herrera, G.; Carillo, M. C. y Torres, A. (2005) "La migración ecuatoriana transnacionalismo, redes e identidades". FLACSO - Sede Ecuador, Vol. 1, 79-108. ISBN, 9978-67-104-8.
- [7] Instituto Nacional de Estadística y Censos. (2010) "Censo de Población y Vivienda". Quito, Ecuador.
- [8] Miles, M. y Huberman, A. (1994) "Qualitative Data Analysis". Estados Unidos: SAGE.
- [9] Organización Internacional para las Migraciones. (2008) "Perfil migratorio del Ecuador". Consultado el 19 de diciembre de 2012, de http://publications.iom.int/bookstore/free/ecuador_profile.pdf.
- [10] Torres, J. E. (2007) "Las condiciones habitacionales y su relación con la pobreza". Departamento Nacional de Planeación. ISBN, 978-958-8340-02-9.
- [11] Universidad Complutense de Madrid. (s.f.) "Análisis Factorial: El procedimiento Análisis factorial". Consultado el 20 de noviembre de 2012, de http://ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/20factor_SPSS.pdf.
- [12] Universidad Complutense de Madrid. (s.f.) "Análisis de conglomerados: El procedimiento Conglomerados jerárquico". Consultado el 25 de octubre de 2012, de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/22conglj.pdf>.
- [13] Universidad Complutense de Madrid. (s.f.) "Análisis de conglomerados: El procedimiento Conglomerados de K medias". Consultado el 30 de octubre de 2012, de http://ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/21conglk_SPSS.pdf.

Analítica

Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación Monte Carlo

Operational Risk Quantification using aggregate loss models and Monte Carlo simulation

Marco Flores



www.inec.gob.ec | www.inec.gob.ec/analitika

Cuantificación del riesgo operacional mediante modelos de pérdidas agregadas y simulación Monte Carlo

Operational Risk Quantification using aggregate loss models and Monte Carlo simulation

Marco Flores

Departamento de Ingeniería Eléctrica y Electrónica, Escuela Politécnica del Ejército, Quito, Ecuador

mjflores@espe.edu.ec

Recibido: 10 de septiembre de 2012

Aceptado: 5 de enero de 2013

Resumen

En este artículo se presenta un software diseñado para estimar la dotación de capital por Riesgo Operacional (RO) utilizando modelos de pérdidas agregadas, siguiendo los requerimientos planteados en Basilea II y utilizando el método Monte Carlo para la solución numérica. Este sistema estima y analiza los parámetros de las funciones de frecuencia y severidad para luego simular la distribución por pérdidas agregadas (LDA), y finalmente calcular la dotación de capital. Para validar la propuesta, se incluyen los resultados de varios experimentos de casos simulados y reales, bajo distintas funciones de distribución clásicas.

Palabras clave: Riesgo operacional, Monte Carlo, distribución de pérdidas, Basilea II, VaR, OpVar, software.

Abstract

This article presents a system to quantify an Operational Risk capital charge, using the LDA method, based on the Basel II regulatory requirements, and the Monte Carlo technique for its numerical solution. Following the Loss Distribution Approach, this system estimates and analyzes the parameters of the frequency and severity distribution models, and finally, it computes the OpVar. Several examples are show to validate the proposed algorithms for measuring the capital charge.

Keywords: Operational risk, Monte Carlo, loss distribution, Basel II, VaR, OpVar, software.

Código JEL: A13 y A30

1 Introducción

El alto porcentaje de pérdidas debido al *riesgo operacional (operativo)* ha influido para que el Comité de Basilea de Supervisión Bancaria implemente medidas técnicas para controlarlo [2]. Varios estudios realizados estiman que las pérdidas atribuidas a este factor se encuentran en alrededor del 35% [2], [4] de las pérdidas totales en las que puede incurrir una institución financiera.

De la definición de este riesgo, “*el riesgo de pérdida resultante de una falta de adecuación a un fallo de los procesos, el personal y los sistemas internos o bien de acontecimientos exter-*

nos” [8], se deduce que varios son los factores que contribuyen a su incremento, entre ellos están procesos internos y externos inadecuados de personas o sistemas, el continuo incremento de los volúmenes de transacciones, la demanda de automatización y velocidad de servicios, etc. [1], [2]. Por ejemplo, el Banco Allied Irish informó de un fraude por USD 691 millones en 2002, que es un caso concreto de pérdidas por RO.

Un adecuado cálculo del RO es beneficioso porque permitiría estimar adecuadamente las pérdidas debido a éste

factor y permitiría tener una mejor valoración del capital de dotación. El Comité de Supervisión Bancaria Basilea II, ha planteado una metodología para evaluar la dotación de capital por RO basada en el enfoque de la distribución por pérdidas agregadas [2], donde se asume que el evento de pérdida es la composición de dos eventos, uno denominado de frecuencia y otro de severidad [16].

Para el cálculo del capital regulatorio se aplica el concepto de Valor en Riesgo (Value at Risk, VaR) al contexto del riesgo operacional, adoptando la nomenclatura de OpVar (Operational Value at Risk). El OpVar representa un percentil de la distribución de pérdidas [8] y refleja grandes pérdidas ocasionadas por el RO. Mientras que el VaR es generalmente definido como el capital suficiente para cubrir pérdidas de un portafolio sobre un período fijo de tiempo. La exposición al riesgo es calculado como un percentil de la función de distribución de pérdidas [10]. Basilea II requiere que el percentil sea del 99.9% [7]. La Figura 1 presenta gráficamente el OpVar, donde se observa que la parte más importante para caracterizar las pérdidas se encuentra en la cola derecha de la distribución.

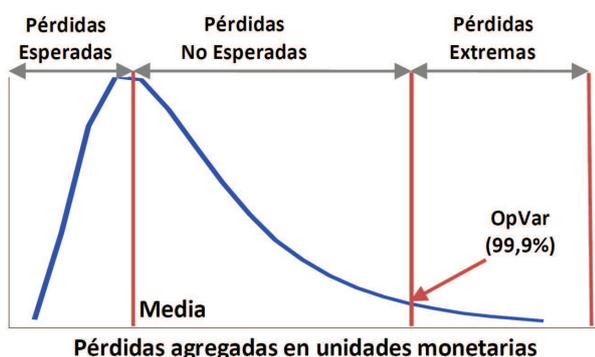


Figura 1. pérdidas agregadas.

Utilizando la metodología de la distribución por pérdidas agregadas, este artículo desarrolla ROMC, un software especializado en la estimación del capital regulatorio por RO. Es así que este trabajo está dividido en cinco partes: La primera, constituida por ésta sección, donde se desarrolla la introducción y motivación de la investigación. En la sección dos se revisa el marco teórico que respalda ésta investigación. Luego, en la sección tres se describe ROMC, el software desarrollado para estimar el OpVar mediante técnicas Monte Carlo. A continuación, en la sección cuatro se presentan varios resultados experimentales con el sistema propuesto. Finalmente, las conclusiones y los trabajos a futuro son presentados en la última sección.

2 Marco teórico

Feria et al. [8] sugiere que la medición se convierte en el aspecto más complejo y, a la vez, más trascendental en el tratamiento del riesgo operacional dentro de las instituciones financieras, debido a que su presencia puede llevar a la quiebra de estas empresas. Por tal motivo, Basilea II propone tres metodologías para calcular los requerimientos de capital por este riesgo, *i*) el método del indicador básico, *ii*) el método estándar y *iii*) las metodologías de medición avanzada (AMA). A su vez dentro de las metodologías AMA se tienen tres metodologías: el modelo de medición interna, los cuadros de mando y el modelo de distribución de pérdidas agregadas [2].

En el enfoque de la distribución por pérdidas agregadas se asume que las pérdidas por RO son la composición de dos eventos, uno denominado de frecuencia y otro de severidad [16]. El primero modela la frecuencia con la que sucede un evento de pérdida en un intervalo de tiempo $[t, t + \delta]$, con $\delta > 0$. Mientras que el segundo, modela la magnitud de dicha pérdida en el mismo horizonte de tiempo.

Consecuentemente, para cuantificar el OpVar, se puede partir del supuesto que la distribución de pérdidas agregadas sigue un proceso estocástico $\{S'_t\}_{t \geq 0}$, formado por dos eventos, uno de pérdida (severidad) y otro de ocurrencia (frecuencia) y relacionados entre si por la ecuación (1), de esta manera, se obtiene el denominado modelo LDA (Loss Data Approach) [1], [3], [4]:

$$S_t = \sum_{k=1} N_t X_k \quad (1)$$

En esta ecuación, $\{X_k\}$ es la magnitud de la pérdida, que generalmente sigue una función de distribución de probabilidad continua. Por otro lado, sea asume que el proceso de conteo N_t sigue una distribución de probabilidad discreta. Además, se supone que estas variables son estadísticamente independientes. Bajo estos supuestos, la media y la varianza vienen dadas por (2) y (3), respectivamente [3].

$$E(S_t) = E(N_t) E(X) \quad (2)$$

$$\text{Var}(S_t) = E(N_t) \text{Var}(X) + \text{Var}(N_t) E(X)^2 \quad (3)$$

Una vez caracterizado el OpVar, es necesario estimar la ecuación (1), para ello, varios métodos han sido propuestos: la recursión de Panjer para estimar numéricamente la distribución de S_t en el caso discreto, la transformación inversa de Fourier [3], simulación Monte-Carlo [12] o la aproximación a la pérdida Simple. La Figura 2 representa gráficamente el proceso de pérdida mediante la composición de los eventos de frecuencia y severidad.



Figura 2. Some neural networks applied in stock market and exchange rate forecasting. Source: author elaboration.

Basilea II ha identificado varias fuentes que generan situaciones de riesgo operacional, y las ha clasificado por líneas de negocios y tipos de riesgos. En el primer caso, se ha identificado 8 líneas de negocios: Banca corporativa, Negocios y ventas, Banca minorista, Banca comercial, Pagos y liquidaciones, Servicios a sucursales, Gestión de activos, y Intermediación minorista. En el segundo caso, se ha identificado siete tipos de pérdidas: Fraude interno, Fraude externo, Prácticas de empleo y seguridad laboral, Clientes, productos y prácticas comerciales, Daños a activos físicos, Interrupción de operaciones y fallos de sistemas, y Ejecución, entrega y gestión de procesos.

De esta manera, se contabilizan 56 fuentes de riesgo operacional, para cada una de las cuales se debe estimar, las distribuciones de frecuencia y severidad y calcular el OpVar. Debido al principio de sub-aditividad del RO, la suma de dos o más riesgos es menor o igual que la suma de los riesgos individuales. Esto genera una acota superior en la estimación del OpVar global [4].

3 Diseño del sistema automático para estimar el OpVar

En la actualidad existen pocos programas autónomos disponibles en el mercado para cuantificar el RO. Generalmente son parte de sistemas mucho mayores como SAS o SPSS [14], por lo tanto, sus costos son elevados y de difícil alcance para instituciones de pequeño o mediano tamaño. Por otra parte, los sistemas de libre distribución como R, presentan dificultades técnicas, porque es necesaria la intervención de personal altamente especializado para desarrollar una metodología.

Pensando en estas dificultades, este sistema ha sido dotado de la suficiente autonomía, automatizando todo el proceso de modelamiento matemático que involucra la evaluación de la ecuación (1). Es así, que se distinguen tres partes: modelamiento de la frecuencia, modelamiento de la severidad y finalmente, estimación del OpVar mediante simulación Monte Carlo.

En este desarrollo se ha utilizado la técnica de simulación Monte-Carlo para estimar la distribución de pérdidas agregadas porque es un método general y directo que se adapta a los distintos casos que pueden surgir en este tipo de distribución. Por otra parte, se han propuesto varios modelos para ajustar las distribuciones de frecuencia y severidad, como se analizará más adelante.

Para construir la primera versión del sistema se han utilizado cuatro funciones para caracterizar la frecuencia y cinco para modelar la severidad. Mediante los estadísticos Chicuadrado y Kolmogorov-Smirnov se seleccionan las mejores distribuciones en cada uno de los casos [11].

Adicionalmente, se debe hacer hincapié en que este software es válido para cuantificar la dotación de capital de cada una de las fuentes de RO identificados en Basilea II [8], y descritas en la sección anterior.

3.1 Moderación de la frecuencia

Para modelar la frecuencia se han utilizado cuatro distribuciones discretas, que en estudios anteriores han demostrado ser las más adecuadas para modelar este tipo de problemas [3], [4]. Sus estimadores se han calculado por el método de momentos, de máxima verosimilitud (MLE) y numéricamente [5], [6]. Los resultados se presentan en la Tabla 1. La distribución Poisson está caracterizada por la tasa de ocurrencia de las pérdidas sobre un periodo constante de tiempo. Una propiedad interesante de esta distribución es que su media y varianza son iguales. Esto implica que si una institución financiera informa de 100 eventos de pérdida en un año, entonces al 95% de confianza se esperaría entre 80 y 120 eventos de pérdida el siguiente año. La distribución Binomial es útil cuando la varianza es menor que la media, viene caracterizada por dos parámetros, uno de los cuales representa el número de riesgos y el otro su probabilidad de suceso. Otra distribución ampliamente usada es la Binomial Negativa, que tiene, al igual que la Binomial, dos parámetros, por lo tanto, es más flexible en su forma que la Poisson. Además, tiene una larga cola, es decir, decae muy ligeramente cuando $r < 1$ y decae más rápidamente cuando $r > 1$.

Tabla 1. Funciones de densidad utilizadas para modelar la frecuencia

Nombre distribución	Densidad $F_N(x)$	Estimación de parámetros
Poisson (λ), $\lambda > 0$, $x = 0, 1, 2, \dots$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$
Binomial (n, p) $p \in (0, 1)$, $n > 1$ $x = 0, 1, \dots, n$	$\binom{n}{x} p^x (1-p)^{n-x}$	Método numérico propuesto por Klugman <i>et al.</i> [3]
Binomial negativa (r, β) $r > 0, \beta > 0$ $x = 0, 1, \dots$	$\binom{x+r-1}{x} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^x$	$\hat{r} = \frac{X^2}{(\text{Var}(X) - \bar{X})}$ $\hat{\beta} = \frac{\text{Var}(x)}{\bar{X}} - 1$
Geométrica (β) $\beta > 0$ $x > 0, 1, 2, \dots$	$\frac{\beta^x}{(1+\beta)^{x+1}}$	$\hat{\beta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

La distribución Geométrica modela el número de fallas que ocurrirían antes de un suceso. Esta función tiene un decaimiento exponencial y la propiedad de pérdida de memoria. Finalmente, en el caso de las distribuciones Binomial Negativa y Geométrica, β está caracterizado por la probabilidad de éxito p , a través de $p = \frac{1}{1+\beta}$.

En los casos Poisson y Geométrica, sus parámetros han sido calculados por el método de máxima verosimilitud [5]. Para la distribución Binomial se ha utilizado un método numérico [3], y en el caso de la Binomial Negativa se ha utilizado el método de momentos [4].

3.2 Elección de la mejor distribución para la frecuencia

Para elegir la mejor distribución que se ajusta a los datos de la frecuencia se ha utilizado el test Chi-cuadrado [11], viene dado por (4):

$$Q = \sum_{k=0}^{\infty} \frac{(n_k - E_k)^2}{E_k} \quad (4)$$

donde E_k es el número esperado de eventos dados por

$$E_k = n \Pr(N = k; \hat{\theta}) \quad (5)$$

y $\hat{\theta}$ representa los parámetros estimados y n_k es el número de eventos en la clase k . Para la implementación computacional se ha seguido las sugerencias planteadas por Klugman *et al* [3], para calcular el número de clases. Este test consiste en medir la mayor distancia entre las distribuciones involucradas. En la práctica se elige la distribución con el menor valor del test, o equivalentemente, el mayor p -valor.

3.3 Modelación de la severidad

Para modelar la severidad se han utilizado cinco distribuciones continuas, las mismas que son ampliamente utilizadas en este caso [3], [4]. Sus parámetros han sido estimados utilizando los métodos de máxima verosimilitud (MLE), de momentos y de percentiles [5], [6], sus resultados se presentan en la Tabla 2.

Las distribuciones LogNormal, Weibull, Rayleigh y Exponencial pertenecen a las denominadas distribuciones de cola ligera (light-tailed). Mientras que la distribución Pareto pertenece a la familia de distribuciones de cola pesada (heavy-tailed).

En los casos LogNormal, Rayleigh y Exponencial, sus parámetros han sido calculados por el método de máxima verosimilitud [5], [6]. Mientras que en el caso Pareto se ha utilizado el método de momentos [5], [4], y para la Weibull se ha implementado el método de percentiles [3].

3.4 Elección de la mejor distribución para la severidad

Para elegir la mejor función de densidad que se ajusta a los datos de severidad se ha utilizado el test de Kolmogorov-Smirnov (KS) [4], [11]

$$D_n = \max |F_n(x) - F(x)| \quad (6)$$

donde D_n es la distancia KS, n es el tamaño de la muestra, $F_n(x)$ es la función de distribución acumulada empírica [15] y $F(x)$ es la función de distribución acumulada, evaluada en los parámetros estimados para el mejor caso. En la práctica, para elegir la mejor distribución se escoge la que corresponde al menor valor de test, o equivalentemente, el mayor p -valor.

Tabla 2. Funciones de densidad utilizadas para modelar la severidad.

Nombre distribución	Densidad $f_N(x)$	Estimación de parámetros
Lognormal (μ, σ^2) , $\mu \in \mathbb{R}, \sigma > 0, x > 0$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ donde $z = \frac{\log(x) - \mu}{\sigma}$	$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\log(x_i) - \hat{\mu})^2$
Pareto (α, θ) , $\alpha > 0$, $\theta > 0, x > 0$	$\frac{\delta\theta^\alpha}{(x + \theta)^{\alpha+1}}$	$\hat{\lambda} = 2 \frac{\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}{\frac{\sum_{i=1}^n x_i^2}{n} - 2 \left(\frac{\sum_{i=1}^n x_i}{n}\right) \left(\frac{\sum_{i=1}^n x_i}{n}\right)}$ $\hat{\theta} = \frac{\left(\frac{\sum_{i=1}^n x_i}{n}\right) \left(\frac{\sum_{i=1}^n x_i^2}{n}\right)}{\frac{\sum_{i=1}^n x_i^2}{n} - 2 \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$
Rayleigh (α) , $\alpha > 0$, $x > 0$	$\frac{x}{\alpha^2} \exp\left(-\frac{x^2}{2\alpha^2}\right)$	$\hat{\alpha} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}}$
Wibull (α, β) , $\alpha > 0, \beta > 0, x > 0$	$\frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$	$\ln(\hat{\beta}) = \frac{c \ln(a) - \log(b)}{c - 1}$ $\hat{\alpha} = -\frac{\ln(\ln(4))}{\ln(b) - \ln(\beta)}$ donde a y b son los percentiles al 25% y 75%, respectivamente; $c = -0,262167$
Exponencial (λ) , $\lambda > 0, x > 0$	$\frac{1}{\lambda} e^{-x/\lambda}$	$\hat{\lambda} = \frac{1}{\left(\frac{1}{n} \sum_{i=1}^n x_i\right)}$

3.5 Cuantificación del OpVar

Este software ha sido diseñado y construido para ser utilizado por *analistas de riesgo* de instituciones financieras, es decir, ha sido desarrollado pensando en que el usuario posee al menos conocimientos básicos de Probabilidades, Estadística y Cálculo, necesarios para interpretar los resultados arrojados por el sistema. Además, se supone que el

usuario dispone de los datos de frecuencia y severidad, información indispensable para calcular el OpVar utilizando el método Monte Carlo.

Este método ha sido propuesto bajo diferentes versiones por distintos autores, como Cruz [4], de la Torre *et al.* [17] y Fontnouvelle [9]; pero en esencia, consiste de los pasos descritos en la Tabla 3.

Tabla 3. Algoritmo para estimar el VaR del RO mediante simulación Monte-Carlo bajo LDA.

- a. Elegir el tamaño de la simulación N .
- b. Elegir las mejores distribuciones para la frecuencia (f_N) y la severidad (f_X).
- c. Simular N variables aleatorias de la distribución f_N .
- d. Para cada de las variables aleatorias de la frecuencia (de magnitud m), simular m variables aleatorias uniformes $U(0, 1)$ y calcular la distribución acumulada inversa, de la función f_X , y sumarlas para obtener una observación de la distribución de pérdida.
- e. Ordenar en orden ascendente los valores generados en (d).
- f. Presentar los resultados.

En este algoritmo es necesario aclarar ciertos pasos. En (a) el tamaño de la simulación N está en estrecha relación con los percentiles de la distribución de pérdida, debido a que estos serán calculados por la fórmula $100 \left(1 - \frac{1}{N}\right) \%$. Generalmente, N toma el valor de 1000, 10000, 100000 o 1000000, que corresponden a los percentiles del 99,9%, 99,99%, 99,999% y 99,9999%, respectivamente. En (e), al ordenar ascendentemente los datos simulados de la distribución de pérdida, se pone en correspondencia las pérdidas esperadas con sus respectivos percentiles de acuerdo al valor de N .

4 Experimentos con ROMC

Varias son las ventajas que presenta este sistema respecto a los existentes en el mercado, las más importantes son: autonomía del programa, estimación simultánea de varias distribuciones para la frecuencia y la severidad, estimación simultánea de los estadísticos Chi-cuadrado y Kolmogorov-Smirnov, cálculo de las estadísticas descriptivas (media, varianza, curtosis y sesgo) para la frecuencia y severidad, tiempos reducidos de cálculo [15], presentación gráfica de los resultados, y finalmente, la capacidad de presentar los resultados de la simulación en formato Word.

4.1 Descripción del sistema

El sistema se encuentra formado por tres pestañas para facilitar su manipulación y uso. En la guía de usuario, que se adjunta al software, se detalla ampliamente los servicios que ofrece. En esta parte, solamente se presentan los resultados experimentales desarrollados para validar el sistema.

4.1.1 Pestaña Sim RO

- i) En esta pestaña se solicita al usuario cargar los datos de frecuencia y severidad, a través de la pantalla de la Figura 3. Los datos deben estar en formato de archivos de texto (txt).

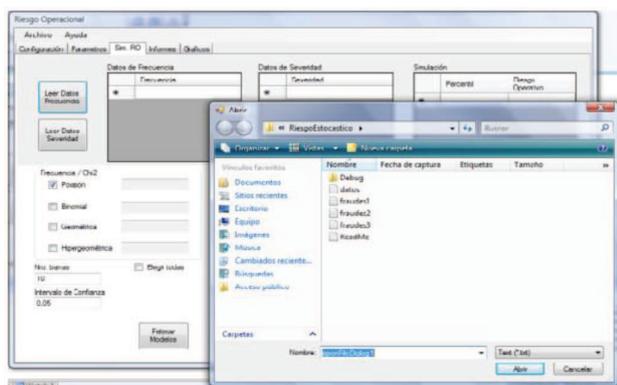


Figura 3. Sistema para estimar el VaR del RO.

- ii) Posteriormente, el usuario puede elegir las distribuciones, tanto de frecuencia como de severidad, o las puede elegir a su criterio y experiencia. En el cálculo se incluyen las estadísticas descriptivas de la frecuencia y la severidad.
- iii) Luego, el usuario puede calcular el OpVar utilizando las distribuciones de frecuencia y severidad, elegidas en el paso anterior, y el algoritmo presentado en la Tabla 3.
- iv) Finalmente, se pueden presentar los resultados en formato Word, de tal manera que se los pueda incluir en otros informes.

4.1.2 Pestaña Gráficos

En esta pestaña se despliega la información gráfica generada por el sistema. En primer lugar, se presenta el histograma de la frecuencia y se contrasta con funciones ajustadas. En segundo lugar, se presenta el histograma de la severidad y se contrasta con funciones ajustadas para el efecto. Finalmente, se presenta el histograma de la distribución de pérdidas agregadas, generada mediante la simulación Monte Carlo.

4.1.3 Pestaña Configuración

En esta pestaña el usuario puede elegir el generador de variables aleatorias de entre 5 opciones: Defecto, Tausworthe, MT19937, Minstd y Knuthran2 [15], [11]. Además, se presenta la información de las estadísticas básicas (media, varianza, sesgo y curtosis) de los datos de la frecuencia y la severidad.

4.2 Resultados experimentales sobre datos simulados

Para desarrollar los experimentos se han simulado los datos de frecuencia y severidad. En el primer caso, se ha generado una muestra de 1000 datos de la distribución Poisson con parámetro $\lambda = 4$. Para la severidad se han utilizado los datos presentado por Cruz [4].

Los resultados se presentan en la Tabla 4, donde se puede observar que las mejores distribuciones son las que tienen el mayor p-valor, es decir, Poisson y LogNormal, respectivamente. En el caso de la Binomial Negativa, sus parámetros no corresponden a los rangos permitidos y por ende, el sistema no los presenta. En este sentido, en la Figura 4 (a) se sobreponen las distribuciones para comparar la calidad de la estimación de la frecuencia, y en la Figura 4 (b) se hace el mismo confrontamiento para la severidad.

Tabla 4. Resultados en el ajuste de la frecuencia y la severidad y elección de las mejores distribuciones sobre datos simulados.

Frecuencia					Severidad				
Distri.	Pmt.		Chi2	p-v	Distri.	Pmt.		KS	p-v
Poisson	$\hat{\lambda}$	4,03	10,54	0,311	LogNormal	$\hat{\mu}$	12,6745	0,1505	0,0603
						$\hat{\sigma}^2$	0,4783		
Binomial	\hat{n}	12	270,48	0,00	Pareto	$\hat{\mu}$	6,1737	0,3132	0,00
	\hat{p}	0,271				$\hat{\theta}$	2275031,977		
Binomial	\hat{r}	-	-	-	Rayleigh	$\hat{\alpha}$	489577,7166	0,4315	0,00
Negativa	$\hat{\beta}$	-							
Geométrica	$\hat{\beta}$	4,03	672,634	0,00	Weibull	$\hat{\alpha}$	1,8226	0,1544	0,05
						$\hat{\beta}$	380225,9146		
					Exponencial	$\hat{\alpha}$	439725,9856	0,2772	0,0

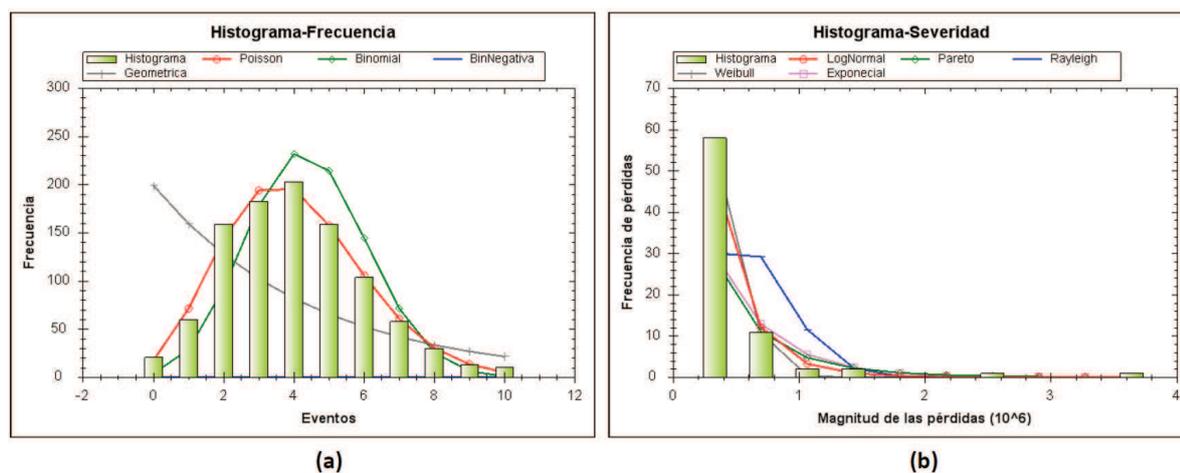


Figura 4. Resultados gráficos del sistema sobre datos simulados, (a) estimación de la frecuencia, (b) estimación de la severidad.

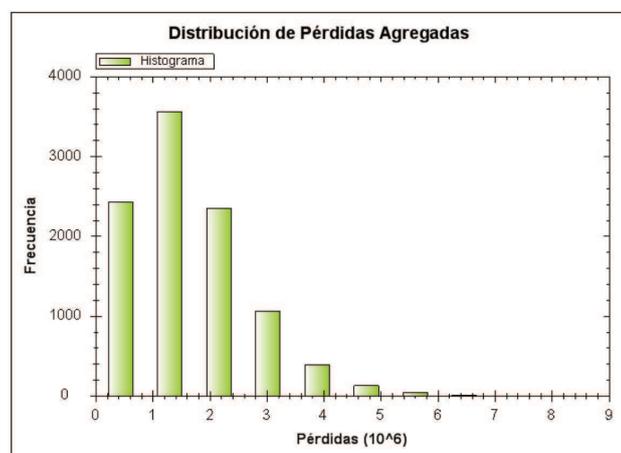


Figura 5. Histograma de la estimación de la distribución de pérdidas para datos simulados.

Para esta simulación, al 99,99 %, se estima que es necesaria una dotación de capital de 8'738.420 UM, y al 99.9 % se calcula que es necesario 6'447.436 UM. En la Figura 5 se observa la distribución de pérdidas obtenida con el módulo gráfico del sistema. Para estas simulaciones se ha elegido la opción MT19937 para generar las variables aleatorias.

4.3 Resultados experimentales sobre datos reales

En este caso, la fuente de datos corresponde a una institución financiera del sector bancario, la cual aporta los datos de pérdidas económicas en los últimos 12 meses, para el sector de préstamos a grandes empresas. Debido a motivos de ética y confidencialidad, el nombre de la institución se omite, y los datos han sido modificados para preservar el anonimato y son medidos en unidades monetarias (UM).

Tabla 5. Resultados en el ajuste de la frecuencia y la severidad y elección de las mejores distribuciones sobre datos simulados.

Frecuencia					Severidad				
Distri.	Pmt.		Chi2	p-v	Distri.	Pmt.		KS	p-v
Poisson	$\hat{\lambda}$	0,8333	29,12	0,0	LogNormal	$\hat{\mu}$	6,7726	0,1469	0,9906
						$\hat{\sigma}^2$	2,7802		
Binomial	\hat{n}	7	32,79	0,00	Pareto	$\hat{\alpha}$	2,7395	0,3555	0,2094
	\hat{p}	0,1389				$\hat{\theta}$	6720,9499		
Binomial Negativa	\hat{r}	-	-	-	Rayleigh	$\hat{\alpha}$	0,8149	0,7244	0,0002
	$\hat{\beta}$	-							
Geométrica	$\hat{\beta}$	0,8333	7,88	0,005	Weibull	$\hat{\alpha}$	0,8149	0,1681	0,9625
						$\hat{\beta}$	1402,2404		
					Exponencial	$\hat{\alpha}$	3863,7475	0,455	0.0497

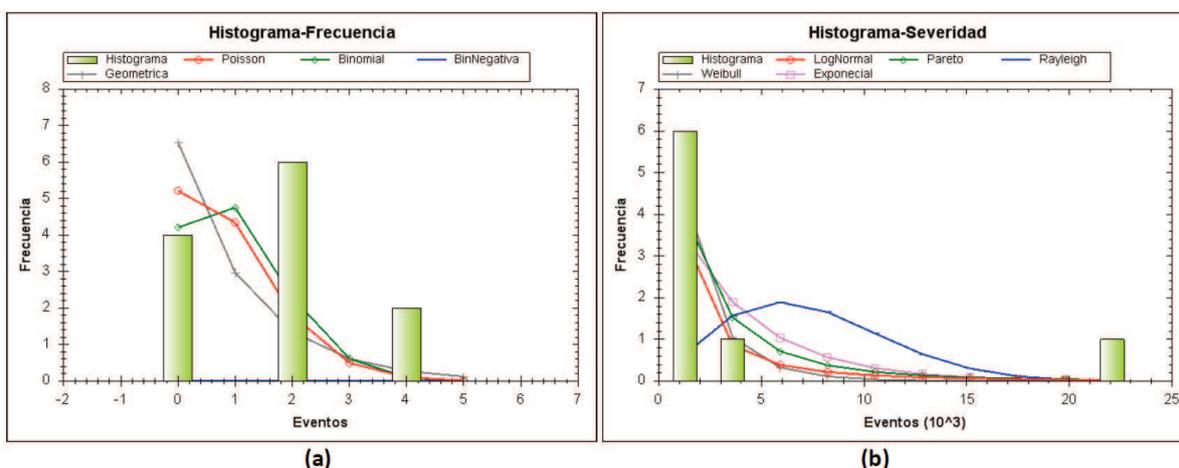


Figura 6. Resultados gráficos del sistema sobre datos reales, (a) estimación de la frecuencia, (b) estimación de la severidad. En ambos casos se presentan los histogramas y las funciones de distribución ajustadas en cada caso.

Los resultados obtenidos se presentan en la Tabla 5, donde se observa claramente que la mejor distribución que explica los datos de frecuencia es la Geométrica, a pesar de que su p -valor es bastante bajo, Figura 6 (a) muestra el histograma y cada una de las funciones ajustadas para este caso. En el caso de la severidad, las distribuciones elegidas son la LogNormal y la Weibull, con p -valores cercanos al uno, la Figura 6 (b) muestra su histograma y cada una de las funciones para este caso.

Una vez elegidas las distribuciones se obtiene los resultados que se presentan en la Tabla 6, el OpVar ha sido calculado al 99.9 % y al 99.99 %. El tiempo promedio de cómputo es de 2500 milisegundos al 99.99 %. Para estas simulaciones se ha elegido la opción MT19937 para generar las variables aleatorias.

Al analizar la Tabla 6 se observa que al utilizar dos distribuciones de severidad, bajo la misma distribución de frecuencia, se obtienen distintos valores del capital de dotación. Por ejemplo, al elegir la distribución Geométrica para la frecuencia, y la LogNormal y Weibull para la severidad, al 99.9 %, se obtienen que la dotación de capital es de 671259 UM y 107165 UM, donde la primera estimación supera en casi 6 veces a la segunda. En la Figura 7 se presenta el histograma de la distribución de pérdidas.

En este escenario, la experiencia del analista es extremadamente importante a la hora de interpretar los resultados. Los test estadísticos por si solos no son suficientes para elegir la frecuencia y la severidad, necesarias para el cálculo de la dotación de capital [9].

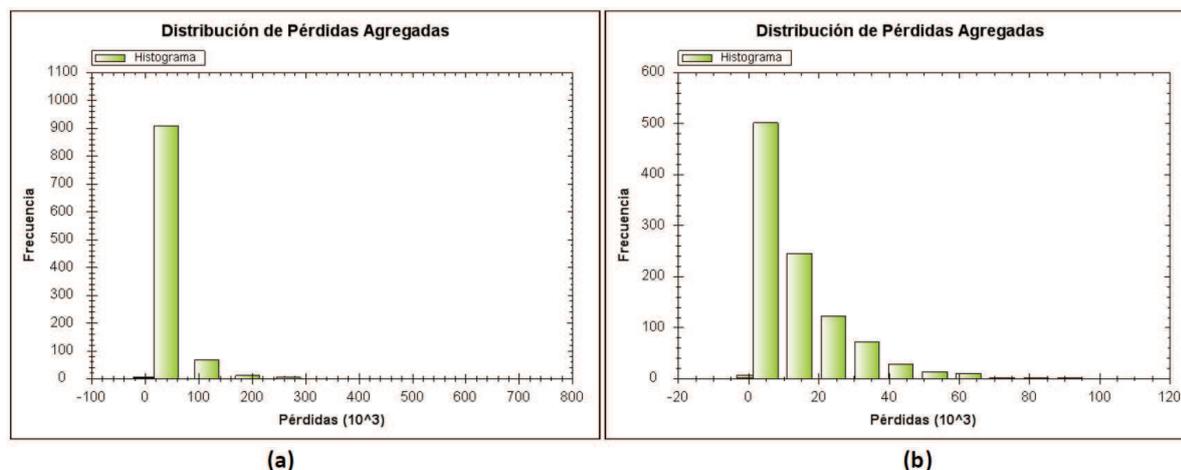


Figura 7. Histograma de la estimación de la distribución de pérdidas, (a) caso Geométrica-LogNormal, (b) caso Geométrica-Weibull.

Tabla 6. Estimación del OpVar al 99.9% y al 99.99%, con varias distribuciones para la frecuencia y la severidad, sobre datos reales.

Simulaciones (N)	Frecuencia	Severidad	Op Var (UM)
1000 (99.9%)	Geométrica	LogNormal	671259
		Weibull	107165
10000 (99.99%)	Geométrica	LogNormal	1459830
		Weibull	139994
1000 (99.9%)	Poisson	LogNormal	216143
		Weibull	18716
10000 (99.99%)	Poisson	LogNormal	573897
		Weibull	25669

5 Conclusiones

En este artículo se ha presentado un sistema automático diseñado para estimar el OpVar utilizando técnicas Monte Carlo bajo el concepto de pérdidas agregadas. Para ello, se han construido varios módulos para analizar los datos provenientes de los eventos de frecuencia y severidad, para posteriormente, estimar la distribución de pérdidas.

En esta primera fase se han utilizado técnicas clásicas de la estadística paramétrica para modelar los eventos de frecuencia y severidad. Una ventaja de este sistema es la automatización en el cálculo de todos los estadísticos necesarios para elegir el mejor modelo en frecuencia y en severidad; además genera a la par todas las simulaciones para calcular el OpVar. Adicionalmente, el sistema presenta un conjunto de gráficos para visualizar todas las distribuciones involucradas en el análisis.

Finalmente, se puede concluir que la interpretación de los resultados depende de la experiencia del analista, y es un tema abierto de investigación que se espera superar en

las próximas versiones del software, incorporando una metodología que ayude en la toma de decisiones.

6 Trabajos a futuro

En el futuro se incorporarán nuevos módulos para ampliar las prestaciones del software. Uno de ellos servirá para estimar los eventos de severidad, utilizando Teoría de Valores Extremos. Asimismo, se incorporará la información interna y externa sobre una estructura bayesiana para mejorar las estimaciones de la frecuencia y la severidad, donde se podrá incorporar, al modelo, la experiencia del analista de riesgos.

Referencias

- [1] Chernobai A., Menn C., Rachev T., and Truck S. (2005): *Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds*, Analysis, pp. 1- 62.
- [2] O.B. Basel Committee Supervision (2001): *Operational Risk*, Distribution.
- [3] Klugman A., Panjer H. and Willmot G. (1998): *Loss Models From Data to Decisions*. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd.
- [4] Cruz, M. G. (2002): *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons, Ltd.
- [5] Casella G., Verger T. (2002): *Statistical Inference*. Thomson Learning.
- [6] Mukhopadhyay N. *Probability and Statistical Inference*. Marcel Dekker, Inc. 2000.
- [7] Dutta K. and Perry J. (2006): *A Tale of Tails :An empirical analysis of Loss Distribution Models for estimating*

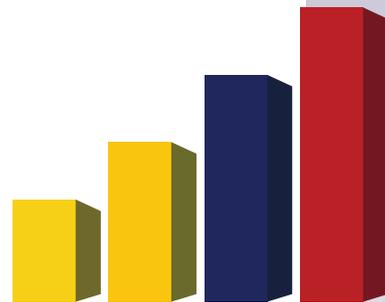
- Operational Risk Capital*. Federal Reserve Bank of Boston.
- [8] Feria J. and Jimenez E. and Martín J. (2007) :*El Modelo de Distribución de Pérdidas Agregadas (LDA), una Aplicación al Riesgo Operacional*. Departamento de Dirección de Empresas, Universidad Pablo de Olavide.
- [9] Fontnouvelle P., Rosengren E. and Jordan J. (2004) :*Implications of Alternative Operational Risk Modeling Techniques*. Federal Reserve Bank of Boston.
- [10] Gilli M. and Kellezi E. (2006) :*An application of Extreme Value Theory for Measuring Financial Risk*, Computational Economics 27(1), pp. 1-23,
- [11] Martinez W. and Martinez A. (2002) :*Computational Statistics Handbook with Matlab*. Chapman & Hall/CRC.
- [12] Lambrigger D., Shevchenko P. and Wuthrich M. (2007) :*The Quantification of Operational Risk using Internal Data, Relevant External Data and Expert Opinions*. Journal of Operational Risk 2(3), pp. 3-27.
- [13] Law A. and Kelton D. (1992) :*Simulation, Modeling and Analysis*. McGraw-Hill Inc..
- [14] SAS System, www.sas.com
- [15] Stroustrup B. (1991) :*The C++ Programming Language*. 2nd edition. Addison Wesley.
- [16] Peters G. and Sisson S. (2006) :*Bayesian Inference, Monte Carlo Sampling and Operational Risk*. Department of Mathematics and Statistics. University of New South Wales, Australia 14
- [17] Torre D. de la, Oris L. Tavecchia D. *Fractal Estimations and Simulations in Operational Risk Analysis*.

Analítica

Cuantificación de las pérdidas inesperadas ocasionadas
por la delincuencia en Ecuador

Quantification of losses caused by delinquency in Ecuador

Yannira Chávez, Patricia Cortez y Paúl Medina



www.inec.gob.ec | www.inec.gob.ec/analitika

Cuantificación de las pérdidas inesperadas ocasionadas por la delincuencia en Ecuador

Quantification of losses caused by delinquency in Ecuador

Yannira Chávez[†], Patricia Cortez[‡] y Paúl Medina^{*}

^{†‡} Instituto Nacional de Estadística y Censos, Quito, Ecuador

^{*} Escuela Politécnica del Ejército, Quito, Ecuador

[†]yanniris.mericita@gmail.com, [‡]patricia.cortez.guevara@gmail.com, ^{*}plmedina@espe.edu.ec

Recibido: 10 de diciembre de 2012

Aceptado: 13 de abril de 2013

Resumen

La presente investigación estudia el impacto económico un conjunto de delitos tipificados en Ecuador como: robo total de vehículo, robo accesorios de vehículo, robo de vivienda, robo a personas, estafa, intimidación o amenaza y heridas. Para cuantificar el costo económico que representa para la sociedad la actividad delictiva desarrollamos un modelo que estima las pérdidas esperadas y las inesperadas, utilizando la metodología Loss Distribution Approach (LDA). En este trabajo, empleamos los datos de la Encuesta de Victimización y Percepción de Inseguridad 2011, realizada por el Instituto Nacional de Estadística y Censos (INEC), y los registros administrativos provenientes de la Dirección General de Operaciones de la Policía Nacional.

Palabras clave: costo, impacto económico, LDA.

Abstract

This paper quantifies the economic cost of several crimes in Ecuador. In particular, we looked at thefts from vehicle, car accessories and homes, robbery of individual, fraud and intimidation (threat and injuries). To do so, we developed a model to estimate expected and unexpected losses, using Loss Distribution Approach (LDA) methodology. We work with data from the Survey of Victimization and Perceptions of Insecurity 2011, carried out by the *National Institute of Statistics and Censuses* (INEC, Spanish acronym) and Administrative Records from the *National Police General Direction of Operations*.

Keywords: cost, economic impact, Loss Distribution Approach.

Código JEL: C15, C63, G20.

1 Introducción

Durante los últimos años se ha notado un incremento en los actos delictivos, pudiendo decirse que constituyen un problema generalizado en la sociedad. Esto conlleva consecuencias físicas, psicológicas y económicas; el impacto de estas últimas es mayor en países en vías en desarrollo, como es el caso del Ecuador. Así, en 2008 según [4] el 12,70 % de personas mayores de 16 años fueron víctimas de al menos uno de los siguientes delitos: robo con fuerza,

robo sin fuerza, ataques y amenazas, es decir, más de una de cada diez personas son o han sido víctimas de ataques delictivos. Para el año 2011 asciende a 17,26 %. También se conoce que el 16,10 % de hogares han sido víctimas de al menos uno de los siguientes delitos: robo a viviendas y/o robo a automóviles, en el transcurso del año 2008, mientras que para el año 2011, se registró el 1,06 % de robo de automóviles, el 10,11 % robos de accesorios, y un 3,57 % robo a

vivienda. Basados en esta breve estadística, no es una sorpresa que los habitantes perciban la inseguridad como un mal social que afecta negativamente a su vida y, por lo tanto, demanden mayor seguridad tanto personal como para sus bienes.

Aún con la información disponible, no es posible determinar la verdadera afectación económica de los delitos. Por un lado, muchos de ellos nunca son denunciados, como confirman los siguientes datos sobre delitos no denunciados: robo con fuerza 76,60 %, robo sin fuerza 95,60 %, robo a viviendas 72,70 %, robo de automóviles 12,80 % y robo de accesorios 85,60 % [4] en el año 2008; en el año 2011, se registran los siguientes porcentajes de delitos no denunciados: robo a personas 84,3 %, robo a vivienda 75,6 %, robo de automóviles 15,5 %, robo de accesorios 81,5 %. Por otro lado, en el caso de los delitos que son denunciados, su afectación económica no necesariamente se corresponde con el valor denunciado.

Para una primera aproximación bastaría con cuantificar el valor establecido en cada uno de los delitos y calcular su media, pero se sabe que este valor no es real pues, sumado a que muchos delitos no se denuncian, está el evento de que muchos otros son de difícil cuantificación, como es el caso de una amenaza. A pesar de todos estos inconvenientes se cree que es posible llegar a estimar un posible valor de pérdida esperada e inesperada, considerando, para ello, el concepto de *Riesgo Operacional*, RO ¹. Cabe indicar que en el presente trabajo, como en el ámbito financiero, cuando se habla de riesgo nos referiremos a la posibilidad de pérdidas causadas por variaciones en los factores que afectan el valor de un activo. En esta investigación, en particular, se trata de evidenciar las pérdidas monetarias a las cuales estaría expuesta la sociedad ecuatoriana, cuando sus activos (vehículos, vivienda, objetos personales) son afectados por los delitos aquí considerados.

La metodología de evaluación del riesgo operacional que aquí utilizaremos será la LDA (Loss Distribution Approach). Un modelo LDA, en riesgo operacional requiere los siguientes supuestos: a) la pérdida total se define como una suma aleatoria de las distintas pérdidas y b) las pérdidas son el resultado de dos diferentes fuentes de aleatoriedad: la frecuencia y la severidad.

Por otra parte, cabe indicar que los datos de *frecuencia* fueron suministrados por los registros administrativos de la Dirección General de Operaciones de la Policía Nacional del Ecuador y los datos de *severidad* se obtuvieron a partir de la Encuesta de Victimización y Percepción de Inseguridad 2011, tomando dos fuentes de información diferentes para garantizar la independencia de la frecuencia y severidad.

El artículo se estructura de la siguiente manera. En la sección 2, se exponen los aspectos teóricos necesarios para el planteamiento del problema. En la sección 3, se presenta los datos, en la sección 4, se presentan los principales re-

sultados del estudio; en la sección 5, se presenta un análisis comparativo y finalmente, en la sección 6, se presentan las principales conclusiones.

2 Marco teórico

En esta sección se plantean los principales aspectos teóricos para analizar la cuantificación de las pérdidas inesperadas por la delincuencia en Ecuador.

El cubrir pérdidas económicas, es decir, medir el capital económico que suman las pérdidas para incluirlo en el balance de la institución, se ha convertido en un aspecto complejo del riesgo operativo, debe garantizarse que, en este parámetro, consten todas las fuentes de riesgo. Por tal motivo, la gestión de riesgos dentro de cualquier institución es necesaria, pues permite tener un mejor manejo y administración en la toma de decisiones en base a las expectativas futuras, pondera las posibilidades de pérdidas esperadas e inesperadas, controla la puesta en práctica de las acciones, y evaluar los resultados de las mismas de manera homogénea y ajustada según el riesgo asumido.

En finanzas, el análisis de cualquier tipo de riesgo cubre tres etapas básicas: identificación, cuantificación y gestión. En la etapa central, la modelación matemática constituye la herramienta primordial. Para el proceso de modelación y cuantificación del riesgo operacional, una de las técnicas más utilizadas es el Método de Distribución de Pérdidas Agregadas, conocido como LDA. Una explicación más detallada sobre las técnicas para la medición de RO se la puede encontrar en [5].

En la Tabla 1, se aprecia que los conceptos en los cuales se basa el método LDA son intercambiables con aquellos usados para calcular el Riesgo Operativo en una Institución Tradicional (sector financiero). Proponemos considerar al Estado, o a la sociedad ecuatoriana, como instituciones no tradicionales, en las cuales es factible aplicar el cálculo de riesgo operacional, con la técnica LDA.

Respecto del intercambio de conceptos, obsérvese que Entidad es reemplazado por Estado; Línea de Negocio, por Delincuencia; y que, en lugar de Tipo de cambio, consideraremos siete eventos de riesgo.

2.1 Método de distribución de pérdidas (LDA)

Los orígenes del LDA se ubican en las aplicaciones actuariales [2], siendo el principal objetivo proporcionar una estimación de pérdida, tanto por línea de negocio como por evento, dicha distribución de pérdida se origina por los eventos de riesgo (frecuencia y severidad) [7].

El modelo LDA proporciona estimaciones para la pérdida agregada; una vez que se cuenta con ese valor, se pasa a calcular la pérdida total. A continuación, se establecen las definiciones:

¹El riesgo operacional es la posibilidad de ocurrencia de pérdidas financieras, originadas por fallas o insuficiencias en los procesos, personas, sistemas internos, tecnología, y en la posibilidad de la ocurrencia de eventos externos imprevistos.

Definición	RO – Institución Tradicional	RO – Institución no tradicional
Entidad	Banco	Estado / sociedad
Línea de Negocio	Inversiones	Delincuencia
Eventos de Riesgo	Tipo de cambio	Robo total de vehículo, robo parcial de vehículo robo de vivienda robo personas, estafa intimidación y heridas

Tabla 1. Conceptos para calcular Riesgo Operativo en Instituciones No Tradicionales (con técnica LDA) y en Instituciones Tradicionales. Fuente: Elaboración propia de los autores.

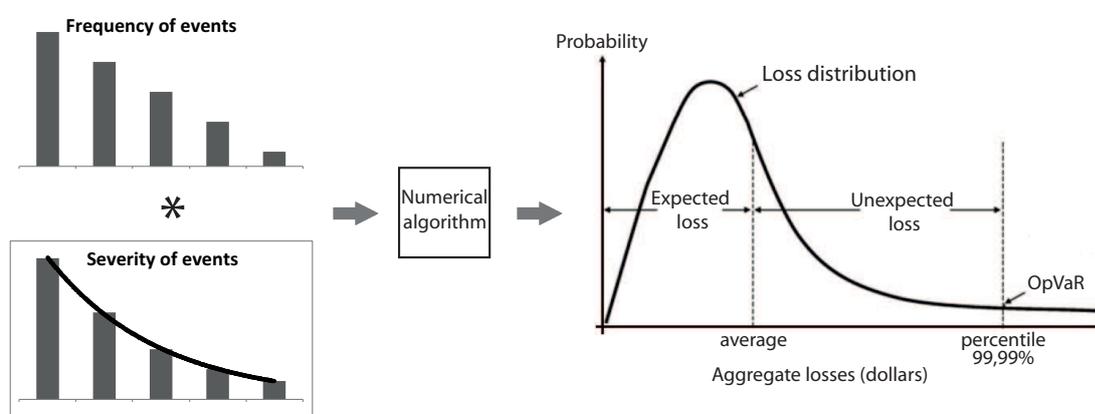


Figura 1. Modelación del LDA. Fuente: Elaboración propia de los autores

DEFINICIÓN 1 (Pérdida total). La pérdida total es la suma aleatoria de las distintas pérdidas.

$$L = \sum_{i=1}^I \sum_{j=1}^J S_{ij}, \quad (1)$$

donde S_{ij} es la pérdida total en la celda i, j de la matriz de pérdidas², I representa el número de eventos considerados, y J el número de líneas de negocio de la entidad. Las S_{ij} se calculan como:

$$S_{ij} = \sum_{N=1}^n X_{Nij}, \quad (2)$$

donde, N_{ij} es la variable aleatoria que representa el número de eventos de riesgo en la celda i, j (frecuencia de los eventos) y X_{Nij} es el monto de la pérdida en la celda i, j (severidad del evento). En consecuencia, las pérdidas son resultado de por lo menos dos diferentes fuentes de aleatoriedad, la frecuencia y la severidad.

Estas estimaciones surgen de combinar, mediante convolución, un proceso estocástico discreto (frecuencia) y un proceso estocástico continuo (severidad). En primer lugar se estima la distribución de frecuencias; luego, la severi-

dad. El siguiente paso es obtener la distribución de las pérdidas totales de riesgo operacional.

Para el círculo de capital regulatorio se aplica el concepto de Valor en Riesgo (VaR por sus siglas en inglés) en el contexto del riesgo operacional, se denomina Valor en Riesgo Operacional (OpVar por sus siglas en inglés).

DEFINICIÓN 2 (Valor en Riesgo Operacional, OpVar). Este representa un percentil de la distribución de probabilidad de pérdidas por lo que es, ante todo, una medición de tipo estadístico.

$$Pr\{L > OpVar\} = p, \text{ donde } p = \text{percentil.}$$

Un esquema de esta técnica se puede ver en la Figura 1. En la Figura 1, es preciso notar que:

- La distribución de frecuencia y severidad se obtienen de la información que se usa para realizar el análisis.
- Las dos distribuciones (frecuencia y severidad) se combinan para obtener la distribución de pérdidas agregadas.
- El OpVar se obtiene tomando el percentil de la distribución de pérdida agregada en el nivel de confianza deseado. Resumiendo, el OpVar se puede interpretar como

²Es la matriz en la que se identifica las líneas de negocio con sus respectivos eventos de riesgo.

una cifra, expresada en unidades monetarias, que nos informan sobre la mínima pérdida potencial en la que podría incurrir una determinada línea de negocio, por tipología de riesgo operacional, dentro de un horizonte temporal de un año y con un nivel de confianza estadístico del 99,9%

- La *pérdida esperada* es la pérdida media que es factible y probable que ocurra en un período determinado. Es decir, las pérdidas esperadas recogerán todas aquellas mermas, previsible y habituales, intrínsecas a la actividad ordinaria de la entidad.
- La *pérdida inesperada* consisten en la diferencia entre el OpVaR y la pérdida esperada, es decir, es el monto de capital que la institución debería tener para cubrir las pérdidas no esperadas a sucesos no previstos inicialmente por la entidad que, sin embargo, pueden desencadenar situaciones funestas para la institución dada la magnitud del quebranto, de acuerdo al nivel de confianza deseado.

La distribución de las pérdidas agregadas es resultado de una composición entre la variable aleatoria discreta asociada a la frecuencia y la variable aleatoria continua asociada a la severidad de los eventos de riesgo.

DEFINICIÓN 3 (Distribución de pérdida Agregada). . *Esta distribución constituye la media ponderada de la n-ésima convolución de la severidad, donde los pesos son las probabilidades de masa de las frecuencias. La n-ésima convolución de la severidad es la probabilidad de ocurrencia del agregado de n pérdidas individuales. Si las pérdidas agregadas para una celda específica están dadas por*

$$S = X_1 + X_2 + \dots + X_N = \sum_{i=1}^N X_i, \quad (3)$$

donde, N es la variable aleatoria de conteo del evento, y X_i es la variable aleatoria de severidad por ocurrencia del evento, donde todas las X_i se asumen independientes y con función de distribución idéntica dada por

$$F_X(x) = P(X \leq x), \quad (4)$$

entonces, la n-ésima convolución de la distribución de severidad, denotada por $F_X^{*n}(x)$, está dada por

$$P(X_1 + X_2 + \dots + X_n \leq x) = F * F * \dots * F = F_X^{*n}(x), \quad (5)$$

por tanto, la función de distribución de las pérdidas agregadas está dada por

$$G_S(x) = P(S \leq x) = \sum_{n=0}^{\infty} P(N = n)F_X^{*n}(x). \quad (6)$$

OBSERVACIÓN 1. El modelo LDA considera los siguientes supuestos dentro de cada clase de riesgo:

- a) la variable frecuencia es una variable aleatoria independiente de la variable aleatoria severidad;
- b) las observaciones de tamaño de pérdidas dentro de una misma clase son homogéneas, independientes e idénticamente distribuidas.

La suposición a) admite que la frecuencia y la severidad son dos fuentes aleatorias e independientes. La suposición b) significa que dos diferentes pérdidas dentro de la misma clase son homogéneas, independientes e idénticamente distribuidas.

En general, resulta muy complejo obtener una solución analítica de la igualdad (6), motivo por lo cual se utilizan métodos numéricos como: simulación Monte Carlo y/o el enfoque recursivo de Panjer. De igual manera, se pueden utilizar métodos analíticos aproximados como la transformada rápida de Fourier, entre otros, (véase e.g. [5]).

Se calculará la distribución de pérdidas agregadas con la simulación Monte Carlo. En el presente documento, este enfoque estima dicha distribución utilizando un número suficiente de escenarios hipotéticos a partir de las distribuciones de severidad y frecuencia, generados aleatoriamente.

2.2 Distribuciones de frecuencia y de severidad

La variable aleatoria *frecuencia* representa el número de eventos que producen pérdidas en un determinado intervalo de tiempo, y que sigue una distribución de probabilidad de referencia. En [5, 9], se muestra que la distribución de Poisson se ajusta a muchas situaciones reales de RO. No obstante, se recomienda considerar otras alternativas, tales como la distribución Binomial Negativa, Binomial, Geométrica e Hipergeométrica.

La variable aleatoria *severidad* representa el impacto del evento en términos de pérdidas económicas y que sigue una distribución de probabilidad de referencia. En [5, 9], se muestra que la distribución Lognormal se ajusta a muchas situaciones reales de RO; se recomienda considerar otras alternativas como la distribución Weibull, Pareto y Exponencial.

3 Datos

El modelo LDA se establece sobre la información de pérdidas históricas registradas en base a la matriz que conforma la línea de negocio (Delincuencia) y a los tipos de riesgo (Delitos). Los datos para trabajar provienen de distintas fuentes de información, ya que se debe satisfacer el supuesto de que la frecuencia y la severidad provienen de dos fuentes independientes de aleatoriedad. La primera es la Encuesta de Victimización y Percepción de Inseguridad

2011; la segunda, son los registros administrativos provenientes de la Dirección General de Operaciones de la Policía Nacional del Ecuador. A partir de la primera fuente, se estimaron los valores de severidad como el valor promedio del perjuicio económico declarado por los encuestados en los correspondientes delitos; con la segunda, es decir, los registros administrativos, se estimaron los valores de frecuencia, se obtuvieron los datos a nivel anual que se muestran en la Tabla 2.

Riesgos	Frecuencia	Severidad
Robo total de vehículo	5 908	\$ 4 825,19
Robo de accesorios de vehículo	5 782	\$ 579,74
Robo de vivienda	13 482	\$ 1 380,77
Robo a personas	21 107	\$ 369,29
Estafa/fraude	10 565	\$ 1 430,75
Intimidación/amenaza	11 331	\$ 443,46
Heridas/lesiones	6 684	\$ 128,65
Total	74 859	\$ 9 157,85

Tabla 2. Datos estimados de frecuencia y severidad obtenidos a partir de la Encuesta de Victimización y de los Registros Administrativos. Fuente: Elaboración propia de los autores.

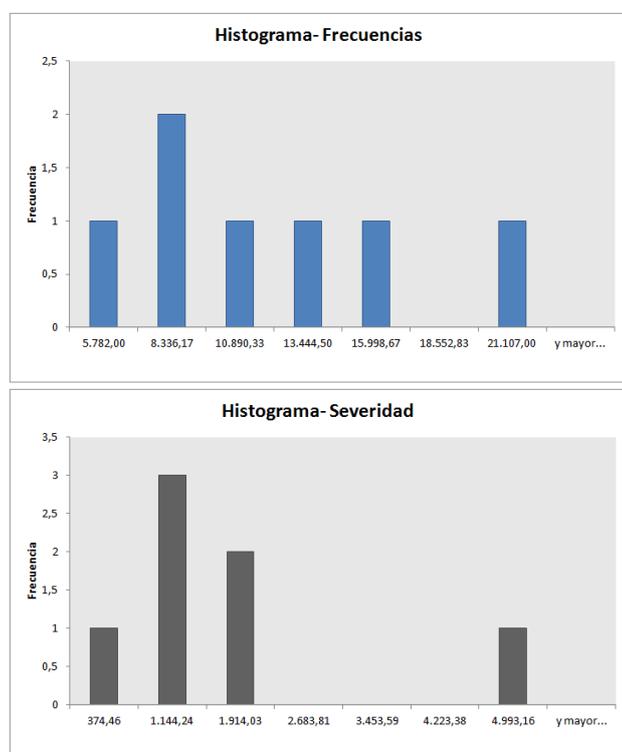


Figura 2. Histogramas de frecuencia y severidad de los delitos. Fuente: Elaboración propia de los autores

En la Tabla 2 se puede observar que la mayor concentración de eventos de riesgo, es decir, mayor frecuencia, es

el robo a personas. El que demuestra mayor severidad, o pérdida económica, es el robo total de vehículo.

En el análisis exploratorio de los datos se puede notar una característica particular sobre la frecuencia de los delitos, mostrando una forma plana de la distribución con un pico; indicando que existe variedad de delitos con una cantidad que suele ser más frecuente sobre los delitos de menor frecuencia (ver Figura 2 arriba). Respecto de la severidad, se puede notar que los delitos con valores bajos presentan mayores severidades que los delitos con valores altos (ver Figura 2 abajo).

Cabe notar, que la información disponible es una limitante en la aplicación de la metodología; técnicamente, resulta complejo encontrar la distribución teórica a la cual se ajustarían los datos observados puesto que, con escasa información los test estadísticos son sensibles, y no cumplen los supuestos establecidos.

4 Estimación de la pérdida por delincuencia

En esta sección se utiliza la metodología y datos descritos anteriormente, se presenta un escenario de la estimación de pérdidas que la delincuencia puede ocasionar en el Ecuador.

Por tanto, desde un punto vista netamente teórico, para estimar estas pérdidas necesitaríamos conocer de manera exacta la función de distribución teórica, discreta o continua, a la cual se pertenecen los valores observados de frecuencia y de severidad, respectivamente. Sin embargo, puesto que la cantidad de datos disponibles para el estudio no permite realizar una inferencia a través de alguna prueba estadística paramétrica o no paramétrica lo suficientemente robusta, no es posible determinar la función de distribución a la cual se ajustarían aquellos datos; por tanto, basados en las recomendaciones de [5, 9], que establecen a Poisson y Binomial Negativa como funciones de distribución típica para la frecuencia, y a Lognormal y Weibull para la severidad, simularemos los cuatro a) Binomial Negativa – Lognormal, b) Binomial Negativa – Weibull, c) Poisson – Lognormal y b) Poisson – Weibull.

Una vez seleccionadas las funciones de distribución a utilizar, se procede a estimar la distribución de pérdidas agregadas (FDP), considerando que

$$FDP = Frecuencia * Severidad.$$

Para realizar esta estimación, se utilizará la metodología de simulación Monte Carlo³. Los resultados obtenidos, considerando las combinaciones citadas, se muestran en la Tabla 3; se realizaron 200000 simulaciones para cada escenario, pues con este número de simulaciones los valores obtenidos tienden a estabilizarse.

³El método de Monte Carlo tiene un error absoluto que decrece como $\frac{1}{\sqrt{N}}$, donde $N \in \mathbb{N}$ es el número de simulaciones a realizar.

Escenarios	Percentil	OpVar	Pérdida Esperada	Pérdida Inesperada
Binomial Negativa y Lognormal	99,90 %	99.968.501.248,68	14.148.719,25	85.819.782,00
	99,00 %	65.476.646.625,25	14.148.719,25	51.327.927,37
	95,00 %	44.355.692.603,31	14.148.719,25	30.206.973,35
	90,00 %	31.563.670.331,79	14.148.719,25	17.414.951,08
Binomial Negativa y Weibull	99,99 %	57.847.808.660,81	13.956.203,02	43.891.605,64
	99,90 %	48.987.077.790,29	13.956.203,02	35.030.874,77
	99,00 %	37.538.486.194,58	13.956.203,02	23.582.283,18
	95,00 %	28.886.900.237,94	13.956.203,02	14.930.697,22
	90,00 %	24.716.681.582,94	13.956.203,02	10.760.478,57
Poisson y Lognormal	99,99 %	93.837.004.281,05	14.150.619,45	79.686.384,83
	99,90 %	61.191.483.458,44	14.150.619,45	47.040.864,01
	99,00 %	39.588.339.684,91	14.150.619,45	25.437.720,24
	95,00 %	28.380.703.759,06	14.150.619,45	14.230.084,31
	90,00 %	23.877.189.441,62	14.150.619,45	9.726.569,99
Poisson y Weibull	99,99 %	46.566.398.141,28	13.957.344,18	32.609.053,96
	99,90 %	38.996.948.828,22	13.957.344,18	25.039.604,65
	99,00 %	31.291.892.402,49	13.957.344,18	17.334.548,22
	95,00 %	25.220.921.006,31	13.957.344,18	11.263.576,83
	90,00 %	22.230.342.740,71	13.957.344,18	8.272.998,56

Tabla 3. Estimación de las pérdidas esperadas e inesperadas, considerando los escenarios para los datos de frecuencia y severidad, respectivamente. Fuente: Elaboración propia de los autores.

En la Tabla 3 se pueden observar las pérdidas inesperadas a diferentes percentiles de cada uno de los escenarios considerados, notando que el cambio de las pérdidas inesperadas es sensible al cambio del percentil. Además, se puede establecer un intervalo de las pérdidas por cada

nivel de percentil, el cual viene dado por las simulaciones Binomial Negativa - Lognormal y Poisson-Weibull.

A continuación, se presenta la distribución total de pérdidas agregadas de los escenarios considerados.

1. Binomial Negativa y Lognormal

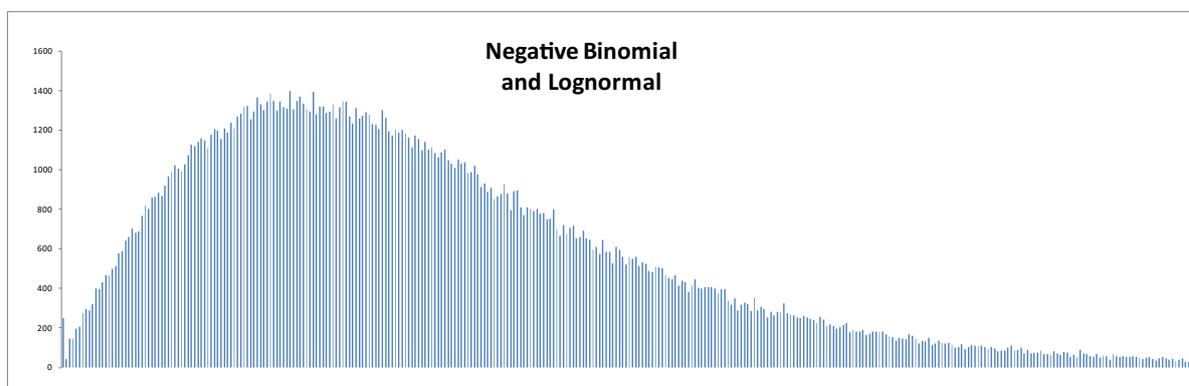


Figura 3. Distribución total de pérdidas agregadas. Fuente: Elaboración propia de los autores.

2. Binomial Negativa y Weibull

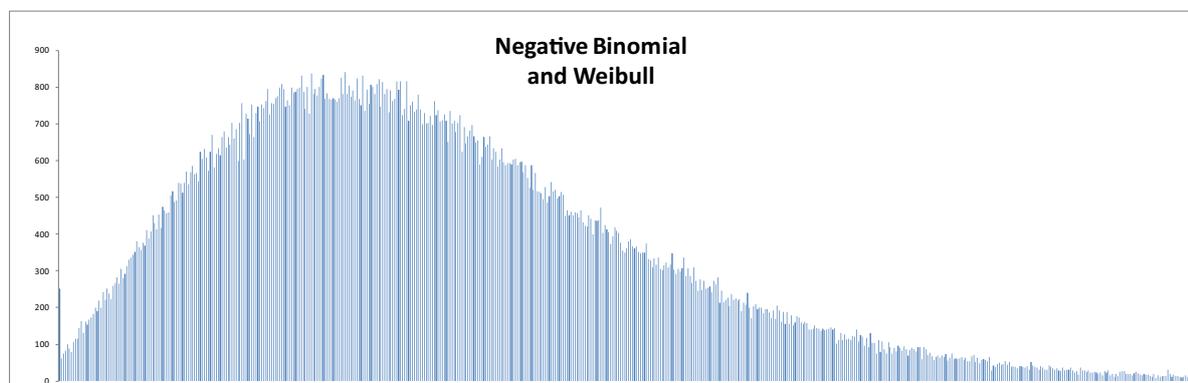


Figura 4. Distribución total de pérdidas agregadas. Fuente: Elaboración propia de los autores.

3. Poisson y Lognormal

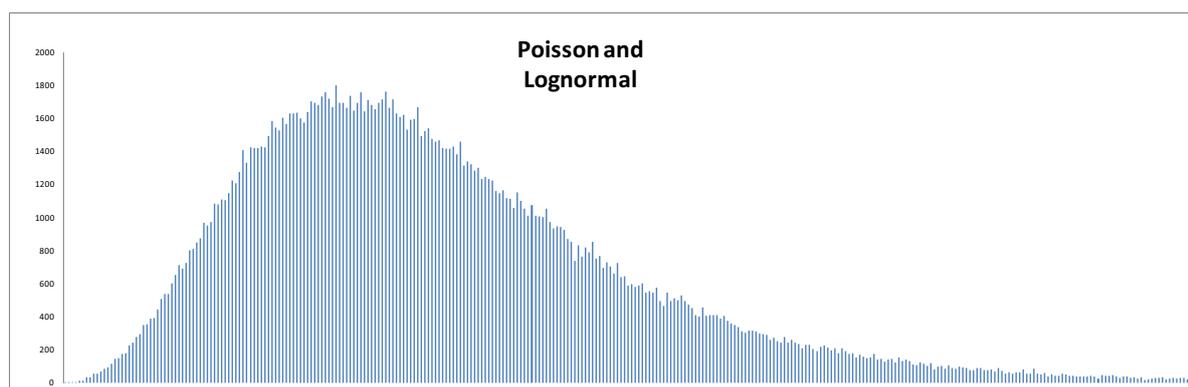


Figura 5. Distribución total de pérdidas agregadas. Fuente: Elaboración propia de los autores.

4. Poisson y Weibull

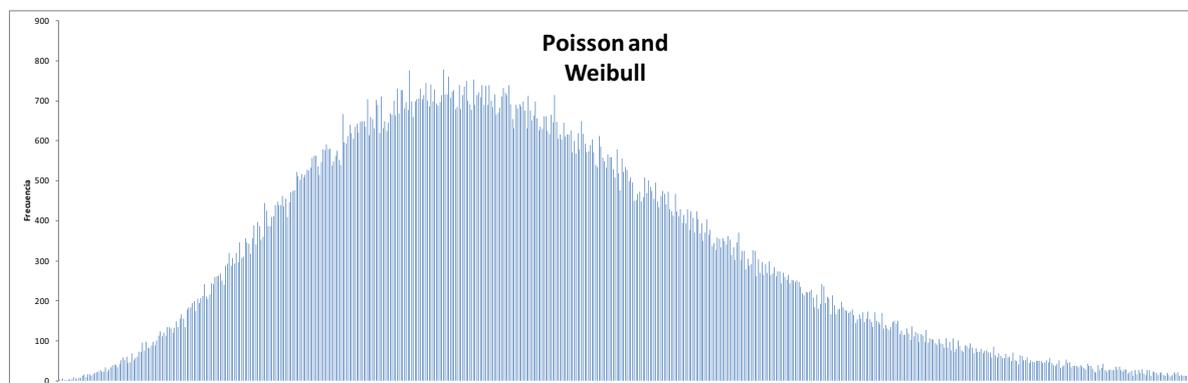


Figura 6. Distribución total de pérdidas agregadas. Fuente: Elaboración propia de los autores.

5 Análisis de resultados

Mediante la estimación de la pérdida por delincuencia en Ecuador, es posible realizar un análisis que permita determinar la sensibilidad de este parámetro al variar la cantidad delictiva; asimismo, es posible también conocer el valor de la pérdida a nivel nacional.

5.1 Análisis de sensibilidad

En esta sección, se realiza un análisis que permite saber, a partir de los resultados obtenidos, en qué porcentaje disminuiría la pérdida al realizar una reducción homogénea en los delitos.

En este sentido, se realiza una disminución consecutiva entera considerando el rango [1, 10], en todos los delitos (frecuencia) analizados, con objeto de conocer la sensibilidad⁴ y variabilidad⁵ que tiene la pérdida ocasionada por la delincuencia.

En referencia al análisis de sensibilidad, en la Tabla 4 y en las Figuras 7, 8, 9 y 10 se muestran los cambios en las pérdidas esperadas que existen tanto de la pérdida esperada como de la pérdida inesperada a diferentes percentiles.

Por un lado, la pérdida esperada en los cuatro escenarios probados se puede notar que la relación que existe

es directamente proporcional y se ve afectada en la misma proporción, es decir que, si se disminuyen a los delitos en 1% la pérdida esperada disminuirá en 1% aproximadamente. Sin embargo, esto ocurre hasta cuando se realiza una afectación del 9%; con una reducción mayor a ésta, se puede observar que la pérdida esperada disminuye, pero en menor proporción, lo cual no es factible.

Por otro lado, no se estabiliza la pérdida inesperada en los cuatro escenarios analizados en los percentiles 99,99%, 99,00% y 95,00% al realizar las reducciones en los delitos; es decir que, si se afecta a los delitos en una proporción X la cola de la distribución de pérdidas agregadas (Ver definición 3) genera ruido afectando a los resultados. Es por esto, que al trabajar con un percentil de 90,00% ya no existe este efecto, y los resultados se estabilizan hasta realizar una afectación del 9%, de igual manera que en la pérdida esperada al realizar un cambio mayor a este los resultados no son factibles.

En resumen, la metodología aplicada para estimar las pérdidas ocasionadas por la delincuencia en el país permite ver los efectos que existirían con una reducción de la frecuencia de los delitos en un valor máximo del 9%; y también permite dar los resultados de la estimación a un nivel de confianza del 90%.

Escenarios	Percentil	Diferencia porcentual con respecto al original									
		1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Binomial Negativa - Lognormal	P.E.	-1,123 %	-1,688 %	-3,272 %	-4,033 %	-5,083 %	-6,270 %	-7,087 %	-8,057 %	-9,007 %	-10,787 %
	99,99%	-2,465 %	3,328 %	-4,778 %	-5,662 %	-7,084 %	-1,617 %	3,023 %	-8,723 %	-10,739 %	-10,882 %
	99,90%	0,048 %	1,924 %	-0,083 %	-1,732 %	-2,202 %	-4,175 %	-1,682 %	-4,978 %	-5,502 %	-1,053 %
	99,00%	-1,288 %	-1,422 %	-2,405 %	-2,382 %	-3,674 %	-5,695 %	-5,310 %	-6,352 %	-5,501 %	2,045 %
	95,00%	-0,649 %	-1,313 %	-2,271 %	-2,318 %	-3,430 %	-4,815 %	-5,699 %	-6,081 %	-6,528 %	1,264 %
90,00%	-0,495 %	-1,126 %	-2,201 %	-2,528 %	-3,753 %	-4,920 %	-5,572 %	-6,378 %	-7,316 %	0,933 %	
Binomial Negativa - Weibull	P.E.	-1,232 %	-1,953 %	-3,229 %	-3,937 %	-5,201 %	-6,018 %	-7,208 %	-8,050 %	-9,139 %	-1,083 %
	99,99%	-1,076 %	1,323 %	-1,302 %	3,329 %	4,283 %	-2,526 %	-5,652 %	-6,258 %	-3,975 %	5,134 %
	99,90%	-2,014 %	-0,688 %	-3,783 %	-2,907 %	-3,614 %	-5,287 %	-6,116 %	-7,057 %	-5,406 %	-1,069 %
	99,00%	-1,237 %	-0,907 %	-2,228 %	-3,351 %	-3,483 %	-4,887 %	-5,250 %	-6,616 %	-6,832 %	1,092 %
	95,00%	-1,246 %	-1,964 %	-2,713 %	-3,559 %	-4,280 %	-5,230 %	-5,990 %	-6,407 %	-7,006 %	0,878 %
90,00%	-0,801 %	-1,715 %	-2,929 %	-3,480 %	-4,433 %	-4,845 %	-6,120 %	-6,271 %	-7,218 %	1,069 %	
Poisson - Lognormal	P.E.	-1,020 %	-2,129 %	-2,971 %	-4,200 %	-5,324 %	-6,267 %	-7,193 %	-8,295 %	-8,861 %	-1,195 %
	99,99%	7,992 %	-11,764 %	2,191 %	-8,548 %	-2,667 %	-1,660 %	6,094 %	-7,004 %	8,220 %	-5,623 %
	99,90%	-0,389 %	-4,199 %	-3,517 %	-2,562 %	-4,286 %	-5,896 %	-6,015 %	-8,467 %	-2,922 %	-1,940 %
	99,00%	-0,539 %	-1,910 %	-1,341 %	-2,051 %	-3,072 %	-3,661 %	-4,176 %	-4,398 %	-3,554 %	-0,248 %
	95,00%	-0,967 %	-1,746 %	-1,189 %	-2,557 %	-3,297 %	-3,162 %	-4,226 %	-4,314 %	-3,691 %	-1,161 %
90,00%	-0,708 %	-1,194 %	-1,213 %	-2,736 %	-3,144 %	-3,624 %	-4,125 %	-4,275 %	-4,806 %	-0,546 %	
Poisson - Weibull	P.E.	-1,203 %	-2,019 %	-3,196 %	-4,205 %	-5,095 %	-6,172 %	-7,242 %	-7,986 %	-9,136 %	-0,847 %
	99,99%	0,053 %	-1,215 %	-5,562 %	-2,989 %	-4,616 %	-4,949 %	-5,918 %	-4,232 %	-3,385 %	-5,408 %
	99,90%	-0,640 %	2,079 %	-1,763 %	-1,818 %	-0,213 %	-3,185 %	-2,551 %	-4,534 %	-4,573 %	-1,447 %
	99,00%	-0,275 %	0,104 %	-1,613 %	-1,801 %	-1,748 %	-2,783 %	-2,833 %	-3,523 %	-3,615 %	-0,472 %
	95,00%	-0,681 %	-0,541 %	-1,498 %	-2,073 %	-2,394 %	-3,072 %	-3,062 %	-3,818 %	-4,288 %	-0,835 %
90,00%	-0,451 %	-0,706 %	-1,274 %	-1,995 %	-2,363 %	-2,636 %	-3,314 %	-3,920 %	-4,213 %	-0,243 %	

Tabla 4. Diferencia porcentual de las variaciones realizadas con respecto al original. Fuente: Elaboración propia de los autores.

⁴Se refiere hasta que punto se puede afectar los delitos.

⁵Se refiere en que porcentaje cambia al variar los delitos.

A continuación, se presenta la afectación por pérdidas inesperadas, a diferentes percentiles.

En la Figura 7 se puede observar claramente que a un percentil de 99,99% existen picos y no se presenta un comportamiento homogéneo al realizar las variaciones; por ejemplo, si la frecuencia de los delitos disminuye en 2% la pérdida inesperada aumenta.

En la Figura 8 se puede observar claramente que a un percentil de 99,99% existen picos y no se presenta un comportamiento homogéneo al realizar las variaciones; por ejemplo, si la frecuencia de los delitos disminuye en 5% la pérdida inesperada aumenta.

En la Figura 9 se puede observar claramente que a un

percentil de 99,99% existen picos y no se presenta un comportamiento homogéneo al realizar las variaciones; por ejemplo, si la frecuencia de los delitos disminuye en 3% la pérdida inesperada aumenta.

En la Figura 10 se puede observar claramente que a un percentil de 99,99% existen picos y no se presenta un comportamiento homogéneo al realizar las variaciones; por ejemplo, si la frecuencia de los delitos disminuye en 4% la pérdida inesperada aumenta.

En referencia al análisis de variación, se muestra en la Tabla 5 en términos monetarios cuál es la variación de la pérdidas esperada e inesperada al realizar una variación en la frecuencia de los delitos de hasta el 9%.

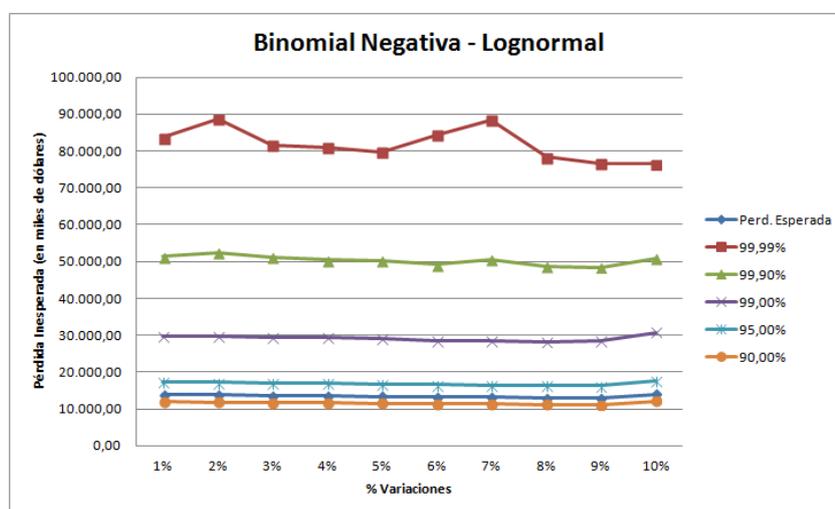


Figura 7. Variaciones por frecuencia a diferentes percentiles. Fuente: Elaboración propia de los autores

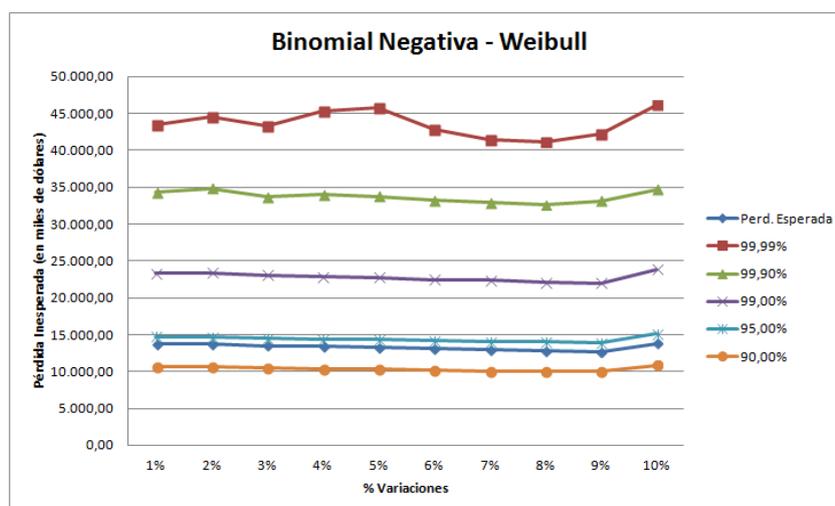


Figura 8. Variaciones por frecuencia a diferentes percentiles.. Fuente: Elaboración propia de los autores

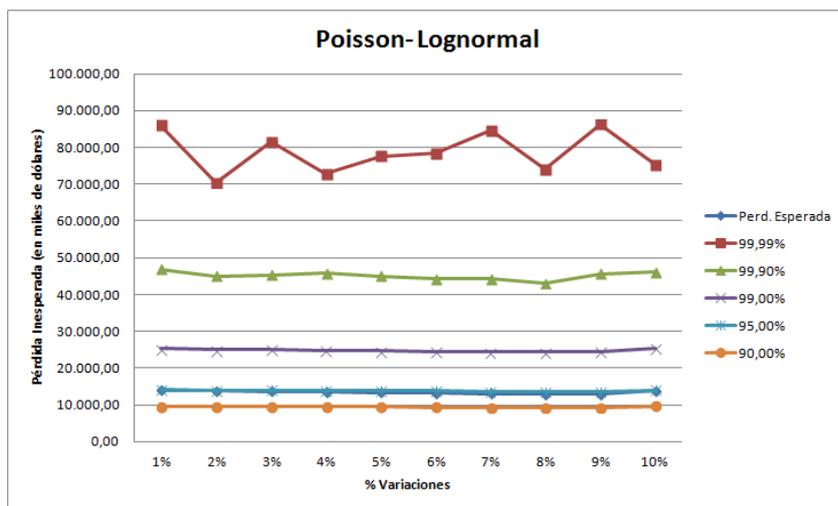


Figura 9. Variaciones por frecuencia a diferentes percentiles. Fuente: Elaboración propia de los autores

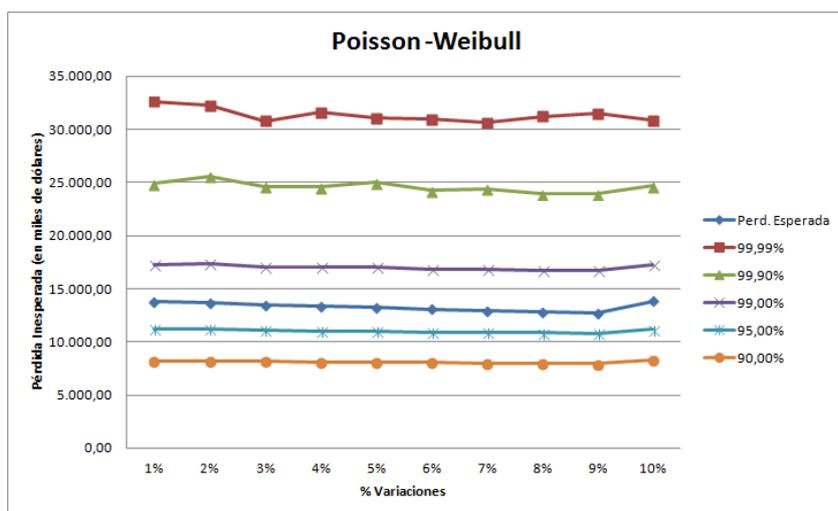


Figura 10. Variaciones por frecuencia a diferentes percentiles. Fuente: Elaboración propia de los autores

Escenarios	Percentil	Original	Variaciones en miles de dólares								
			1 %	2 %	3 %	4 %	5 %	6 %	7 %	8 %	9 %
Binomial Negativa - Lognormal	PE.	14.148,72	13.989,83	13.909,94	13.685,74	13.578,11	13.429,54	13.261,62	13.145,99	13.008,76	12.874,35
	99,99 %	85.819,78	83.704,07	88.675,99	81.719,33	80.960,68	79.740,42	84.431,81	88.414,38	78.333,35	76.603,97
	99,90 %	51.327,93	51.352,61	52.315,67	51.285,14	50.438,80	50.197,50	49.185,12	50.464,39	48.772,70	48.503,75
	99,00 %	30.206,97	29.817,96	29.777,53	29.480,51	29.487,52	29.097,21	28.486,58	28.603,00	28.288,30	28.545,25
	95,00 %	17.414,95	17.301,88	17.186,27	17.019,54	17.011,24	16.817,60	16.576,35	16.422,42	16.356,01	16.278,02
	90,00 %	12.010,04	11.950,64	11.874,75	11.745,67	11.706,42	11.559,25	11.419,20	11.340,80	11.244,07	11.131,34
Binomial Negativa - Weibull	PE.	13.956,20	13.784,29	13.683,59	13.505,57	13.406,69	13.230,33	13.116,31	12.950,18	12.832,71	12.680,78
	99,99 %	43.891,61	43.419,47	44.472,37	43.319,93	45.352,87	45.771,47	42.783,10	41.410,75	41.144,66	42.147,09
	99,90 %	35.030,87	34.325,36	34.789,92	33.705,68	34.012,70	33.764,72	33.178,85	32.888,54	32.558,59	33.137,13
	99,00 %	23.582,28	23.290,68	23.368,43	23.056,95	22.792,08	22.760,98	22.429,87	22.344,18	22.022,12	21.971,17
	95,00 %	14.930,70	14.744,63	14.637,42	14.525,70	14.399,31	14.291,64	14.149,86	14.036,34	13.974,10	13.884,66
	90,00 %	10.760,48	10.674,30	10.575,99	10.445,30	10.386,01	10.283,46	10.239,13	10.101,89	10.085,69	9.983,82
Poisson - Lognormal	PE.	14.150,62	14.006,35	13.849,39	13.730,27	13.556,27	13.397,25	13.263,80	13.132,74	12.976,78	12.896,73
	99,99 %	79.686,38	86.055,04	70.312,18	81.432,13	72.874,45	77.560,76	78.363,29	84.542,12	74.105,15	86.237,00
	99,90 %	47.040,86	46.857,82	45.065,45	45.386,64	45.835,58	45.024,83	44.267,31	44.211,34	43.057,72	45.666,32
	99,00 %	25.437,72	25.300,50	24.951,78	25.096,71	24.915,96	24.656,31	24.506,38	24.375,53	24.318,98	24.533,72
	95,00 %	14.230,08	14.092,48	13.981,65	14.060,96	13.866,25	13.760,86	13.780,18	13.628,74	13.616,20	13.704,92
	90,00 %	9.726,57	9.657,73	9.610,47	9.608,61	9.460,50	9.420,77	9.374,13	9.325,35	9.310,80	9.259,13
Poisson - Weibull	PE.	13.957,34	13.789,50	13.675,50	13.511,27	13.370,37	13.246,26	13.095,89	12.946,59	12.842,64	12.682,22
	99,99 %	32.609,05	32.626,42	32.212,82	30.795,47	31.634,26	31.103,81	30.995,23	30.679,21	31.229,04	31.505,40
	99,90 %	25.039,60	24.879,26	25.560,07	24.598,24	24.584,45	24.986,15	24.242,08	24.400,88	23.904,42	23.894,65
	99,00 %	17.334,55	17.286,83	17.352,62	17.054,88	17.022,31	17.031,62	16.852,10	16.843,37	16.723,77	16.707,92
	95,00 %	11.263,58	11.186,85	11.202,60	11.094,89	11.030,13	10.993,93	10.917,59	10.918,70	10.833,57	10.780,58
	90,00 %	8.273,00	8.235,65	8.214,59	8.167,56	8.107,93	8.077,53	8.054,90	7.998,82	7.948,66	7.924,50

Tabla 5. Estimación de las pérdidas esperadas e inesperadas (en miles de dólares) al realizar variaciones en la frecuencia de los delitos. Fuente: Elaboración propia de los autores.

En la Tabla 5, considerando el percentil 90,00 %, se puede observar en los escenarios que al reducir un 9 % , en la frecuencia de los delitos, la pérdida inesperada se ve afectada de la siguiente manera:

- Binomial Negativa - Lognormal: la pérdida inesperada pasa de 12,01 millones a 11,13 millones.
- Binomial Negativa - Weibull: la pérdida inesperada pasa de 10,76 millones a 9,98 millones.
- Poisson - Lognormal: la pérdida inesperada pasa de 9,72 millones a 9,25 millones.
- Poisson - Weibull: la pérdida inesperada pasa de 8,27 millones a 7,92 millones.

Además, se puede notar que el nivel de precisión aumenta al bajar el nivel de confianza; el intervalo está dado entre Binomial Negativa-Lognormal y Poisson -Weibull.

5.2 Comparación de nivel de pérdidas

Desagregación	PEA
Nacional	6.581.621
Quito	756.923
Guayaquil	1.191.204
Cuenca	199.989
Machala	121.464
Ambato	106.648

Tabla 6. Población Económicamente Activa a partir de los 15 años, a nivel nacional y en las 5 principales ciudades, en 2011. Fuente: Instituto Nacional de Estadística y Censos.

Con la finalidad de saber cuánto representaría la pérdida inesperada por persona económicamente activa, se realiza una comparación entre la pérdida inesperada estimada y la Población Económicamente Activa (PEA) del año 2011.

De la Tabla 6, se toma el valor de la Población Económicamente Activa a nivel nacional para el año 2011, y se realiza lo siguiente:

$$AE = \frac{PI}{PEA} \quad (7)$$

donde, *AE* corresponde a la afectación económica por persona económicamente activa, *PI* es la pérdida inesperada ocasionada por la delincuencia, *PEA* es la Población Económicamente Activa. De esta manera, se obtiene una estimación de la pérdida anual a nivel nacional por persona económicamente activa, siendo:

Escenarios	Percentil	
	90 %	99,99 %
Escenario 1	\$1,825	\$13,039
Escenario 2	\$1,635	\$6,669
Escenario 3	\$1,478	\$12,107
Escenario 4	\$1,257	\$4,955

Tabla 7. Afectación Económica en 4 escenarios. Fuente: Elaboración propia de los autores.

Escenario 1: Binomial Negativa -Lognormal. Escenario 2: Binomial Negativa- Weibull. Escenario 3: Poisson-Lognormal. Escenario 4: Poisson-Weibull.

En la Tabla 7, se puede observar que a un nivel de confianza del 90 % se tiene una pérdida entre \$1,25 a \$1,82 por

persona económicamente activa a nivel nacional. Si el valor obtenido por persona, se lo traslada a nivel de las cinco ciudades principales se obtienen los siguientes resultados, a un nivel de confianza del 90 % y tomando en cuenta los cuatro escenarios.

En la tabla 8 se puede ver que en la ciudad de Guayaquil se presentaría la mayor pérdida encontrándose en un intervalo de 1,4 millones a 2,2 millones de dólares, le sigue la ciudad de Quito, cuyas pérdidas se encuentran en un intervalo de 951 mil a 1,3 millones de dólares.

Ciudades	Escenario 1	Escenario 2	Escenario 3	Escenario 4
Quito	\$ 1.381.220,52	\$ 1.237.514,03	\$ 1.118.608,88	\$ 951.440,19
Guayaquil	\$ 2.173.690,72	\$ 1.947.533,15	\$ 1.760.406,61	\$ 1.497.325,50
Cuenca	\$ 364.936,48	\$ 326.967,35	\$ 295.551,06	\$ 251.382,91
Machala	\$ 221.644,72	\$ 198.584,12	\$ 179.503,38	\$ 152.677,79
Ambato	\$ 194.608,72	\$ 174.361,02	\$ 157.607,73	\$ 134.054,30

Tabla 8. Afectación económica en las cinco principales ciudades del Ecuador, considerando los cuatro escenarios. Fuente: Elaboración propia de los autores.

6 Conclusiones

Para estimar las pérdidas causadas por un conjunto de delitos en Ecuador, se ha utilizado un modelo de valoración de pérdidas generalmente aplicado en estudios actuariales en valoración de riesgo operativo. Se han obtenido diferentes niveles de pérdida, esperada e inesperada, considerando distintos percentiles para cada uno de los cuatro escenarios presentados.

Tomando casi la totalidad de la distribución de pérdidas, es decir, el percentil 99,99 %, se puede llegar a establecer un intervalo de pérdidas causadas por los delitos considerados. El intervalo se establecería entre los \$33 y \$86 millones, aproximadamente. Estos valores vendrían dados por los resultados de las simulaciones Poisson – Weibull y Binomial Negativa – Lognormal, respectivamente.

Es posible observar la diferencia sustancial que se establece entre los valores de pérdida al cambiar el nivel del percentil que se considera para establecer éstas. Así, si se toma el percentil de 99,00 %, el intervalo de pérdidas que se puede establecer sería de \$17 y \$33 millones, aproximadamente.

Se ha establecido que la metodología permite ver cambios en la estimación de la pérdida a un nivel de confianza del 90 % y realizando reducciones de hasta el 9 %.

Se ha determinado que al considerar que las personas económicamente sufran al menos uno de los delitos estudiados se tendría un valor aproximado de pérdida entre 1,25 y 1,85 dólares anuales.

Finalmente, es preciso notar que, al realizar mayores cambios en la reducción de la frecuencia de los delitos, los resultados no son estables; esto puede deberse a la limitación de los datos usados.

Referencias

[1] Aue, F. y Kalkbrener, M. (2006). *LDA at work: Deutsche Bank's approach to quantifying operational risk*, Journal of

Operational Risk, 1, 88-129.

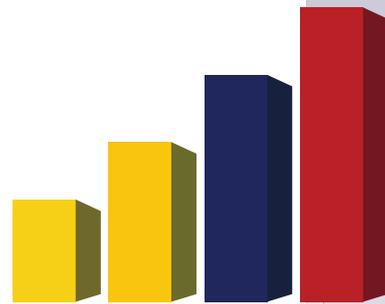
- [2] Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*, Springer-Verlag, Heidelberg, 45-93.
- [3] Cajas J. (2011). *Modelos de enfoque de Medición Avanzado de Riesgo Operativo (EMA)*, Contribuciones a la Economía, Grupo Eumed.net, Universidad de Málaga.
- [4] Capa, H. y Gallardo C. (2008). *Encuesta de Victimización y Percepción*, 1-109.
- [5] Cruz, M. (2002). *Modeling, measuring and hedging operational risk*, New York: John Wiley & Sons.
- [6] Franco, L. y Murillo, J. (2008). *Loss Distribution Approach (LDA): Metodología actuarial aplicada al riesgo operacional*, Revista Ingenierías Universidad de Medellín, 7, 143-156.
- [7] Murillo, J. (2009). *Cuantificación de las pérdidas económicas por riesgo operacional asociadas a eventos extremos*, Universidad Nacional de Colombia, 2-63.
- [8] Nieto, M. (2004). *The modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee*, Documento de trabajo, Banco de Italia.
- [9] Shevchenko, P. y Wéthrich, M. (2006). *The structural modeling of operational risk via Bayesian inference: combining loss data with expert opinions*, Journal of Operational Risk, 1, 3-26.
- [10] Claramunt, M. Boj, E. Costa, T. y Mármol, T. (2012). *Sistemas Informáticos aplicados a la matemática actuarial no vida una propuesta con R*, Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA, 13, 1-26.

Analítica

Normas para la presentación de trabajos

Norms for submission of papers

Código de ética



www.inec.gob.ec | www.inec.gob.ec/analitika

Normas para la presentación de trabajos

Los autores están invitados a enviar sus artículos a la revista Analítica, ya sea en español o en inglés.

Para la publicación de los artículos, los autores deben seguir los siguientes lineamientos:

Originalidad

El trabajo presentado se entenderá que es inédito, original y que no se ha presentado a ningún otro medio antes de conocer la decisión de esta revista. Si el artículo ha sido publicado en otra revista, impresa o digital, este deberá tener el permiso o autorización de la revista, detallando el nuevo aporte o modificación correspondiente. Después de la aceptación, la autorización deberá ser enviada al editor de la revista Analítica.

Presentación

Los artículos deberán ser presentados en formato digital, el cual puede ser *.pdf* o *.doc*. Comprometiéndose el autor o los autores, a proporcionar el trabajo en formato *.tex* (siguiendo los estándares de la American Mathematical Society (AMS)), si el mismo es aceptado para su publicación.

Toda la documentación correspondiente debe remitirse al editor de la revista Analítica, al siguiente correo electrónico: analitika@inec.gob.ec / analitika@analitika.ec.

Requerimientos técnicos

Para ser sometidos a revisión, los trabajos deberán cumplir los siguientes requerimientos técnicos:

- Las ilustraciones y tablas, deben estar numeradas secuencialmente, incluyendo una descripción explicativa para cada una.
- Las figuras deben tener formato *.eps*, de preferencia, o formato *.jpg*, o *.png* de alta resolución.
- El texto en las figuras y mapas debe escribirse con letras fácilmente legibles.
- Diferentes elementos de un gráfico deben contener sus correspondientes explicaciones. Por ejemplo, en un gráfico de pastel se deberá especificar qué representa cada color.

- Asegúrese de que las inscripciones o detalles, así como las líneas, tengan tamaños y grosores adecuados; de tal manera que sean legibles (números, letras y símbolos).
- Las fotografías deben grabarse con alto contraste y en alta resolución. Recuerde que las fotografías frecuentemente pierden contraste en el proceso de la impresión. No combine fotografías y dibujos en la misma figura.
- Si las figuras, gráficos, ilustraciones o fotografías incluidas en el trabajo tienen derecho de autor, es responsabilidad del autor que presenta el trabajo a evaluación, obtener el permiso correspondiente para su uso.
- Los gráficos, ilustraciones y fotografías incluidas en el trabajo deben ser enviados en una carpeta exclusiva, la cual debe llevar el nombre: "Figuras_(las dos primeras iniciales del título del artículo en mayúsculas)".
Por ejemplo, si el trabajo tiene como título: "Densidad poblacional del Ecuador continental", el nombre de la carpeta debe ser: "Figuras_DP".
- El formato general para la etiquetas identificativas de los diversos ambientes deberá ser:

Para tablas:
`\label{DP_tab:001}`

Para figuras
`\label{DP_fig:001}`

Para ecuaciones
`\label{DP_eq:001}`

Para definiciones
`\label{DP_def:001}`

Para teoremas
`\label{DP_teo:001}`

etc.

Estructura y formato

La estructura del documento será de la siguiente forma:

Obligatorias	Opcionales
Título	Agradecimientos
Resumen	Recomendaciones
Palabras clave	Futuros trabajos
Abstract	Anexos
Keys words	
Clasificación	
Introducción	
Marco Teórico o Metodología	
Aplicación o Resultados	
Conclusiones	
Referencias	

Tabla 1. Estructura del documento a ser evaluado.

Formato del documento

Detallaremos las particularidades a considerar, para algunos de los componentes del documento.

Título. El título debe incluir la siguiente información: título del artículo, nombre completo del autor o de los autores, dirección del autor o de los autores, incluyendo su correo electrónico.

Resumen (Abstract). El resumen debe ser en español y en inglés. El resumen debe ser corto y conciso (máximo 250 palabras) y en el mismo se debe expresar los resultados relevantes del estudio. Debe ser informativo y no indicativo; por ejemplo diga:

“En este trabajo se analiza el papel que desempeñan las características individuales, familiares y laborales en la probabilidad de tener un contrato temporal frente a un contrato indefinido,”

lo cual es informativo. No diga

“La probabilidad de tener un contrato temporal frente a un contrato indefinido es discutido y aceptado,”

lo cual es indicativo.

Palabras claves (Keywords). Las palabras claves deben ser de tres a seis y, representarán los principales temas del artículo. Deberán ser colocadas al final del resumen y del abstract, respectivamente.

Clasificación. Se debe incluir el sistema de clasificación del Journal of Economic Literature, JEL.

Referencias. La sección de referencias debe incluir todas las publicaciones citadas en el texto. No se debe incluir reportes no publicados u otro tipo de información que no es posible verificar. Las tesis deben citarse únicamente cuando estén disponibles para consulta en una biblioteca física o virtual. En el texto las referencias deben ir en orden alfabético y deben seguir el sistema “inicial del nombre apellido”. Las referencias de portales electrónicos deben seguir un esquema similar al de las publicaciones, pero adicionalmente se deberá incluir la fecha de consulta. A continuación se muestra un ejemplo:

Referencias

- [1] M. AGUAYO, y E. LORE, *Cómo hacer una Regresión Logística binaria paso a paso II análisis multivariante*, Fundación Andalucía Beturia para la Investigación en Salud, Dot. N° 0702013, (2007).
- [2] C. GAMERO, *Satisfacción Laboral y tipo de contrato en España*, Investigative Radiology 34, Vol. 10, 636 - 642, (1999).
- [3] LEY No. 100, *Código de la niñez y adolescencia*, Registro Oficial 737, Enero, (2003).
- [4] SCIENTIFIC INSTRUMENT SERVICES, Inc. 2006. *MASS Spectral Library*. En línea: <http://www.sisweb.com/software/ms/nist.htm>, Consulta: 10 de abril del 2010.

Agradecimientos. Los agradecimientos podrán incluirse como una sección, la cual deberá estar al final del artículo y constará de un párrafo.

Procedimiento editorial

Cuando el trabajo ha sido recibido por el editor, éste pasa por dos procesos de calificación. El primero, se lleva a cabo dentro de un consejo editorial interno que determinará la pertinencia y solvencia científica y, el segundo, es la evaluación por expertos nacionales o extranjeros, considerando el método *blind review*.

Si el artículo es aceptado para su publicación, éste pasará por las siguientes etapas:

1. El editor enviará al autor principal las observaciones realizadas por los evaluadores para que éste realice las correcciones y cambios necesarios. El tiempo asignado para esto no será mayor a un mes.
2. Cuando el trabajo corregido es regresado al editor, éste hará la copia-edición y empezará el proceso de formato. Después de que el trabajo ha sido formateado para la publicación, una prueba de impresión (“page proofs”) será enviada al autor para corregir posibles

errores. En este punto, no será posible hacer cambios en el documento, sino solamente corregir errores de edición o formato. El autor debe retornar la prueba de impresión al editor con sus observaciones, si las hubiere, máximo en una semana después de haberla recibido.

3. Finalmente, el documento corregido será archivado hasta que la revista empiece su proceso de impresión. Cuando esto ocurra, se enviará una copia impresa y una digital de la revista al autor.

Si el artículo no es aceptado para su publicación, será devuelto al autor notificando el motivo.

Comentarios finales

La publicación del volumen de la revista depende de la colaboración entre los autores, los revisores, la imprenta y el editor. La colaboración y cumplimiento de los plazos establecidos es fundamental.

La revista Analítica publica dos volúmenes por año.

Norms for submission of papers

All authors are invited to send their studies to Analítica Journal in Spanish or English; for being eligible to publish, the paper must meet the following guidelines:

Originality

The paper must be original, unpublished and not be submitted to any other media before knowing the decision of this journal. If the paper has been previously published in another journal, printed or digital, it must have a written permit or authorization detailing any addition or modification made since the last publication. After the acceptance of the paper, such authorization must be sent to the Editor of Analítica Journal.

Format

The papers must be sent in digital format, either **PDF** or **DOC**. If accepted for publication, the author(s) must send the document in **TEX** (according to the standards of the American Mathematical Society (AMS)) format.

All the documentation must be emailed to the Editor of Analítica Journal to: analitika@inec.gob.ec / analitika@analitika.ec.

Technical Requirements

For being subject for review, a paper must meet the following technical requirements:

- Each illustration and table must include an explicative description and be sequentially numbered.
- Pictures must be in format EPS (preferable), JPG or high-resolution PNG.
- Text inside maps or pictures must be easily legible.
- Each element within a graphic must contain its own explanation. For example, a pie chart must show what each color represents.
- Make sure the inscriptions, details and lines have adequate sizes and dimensions, so all numbers, characters and symbols are legible.

- Photographs must be saved in high contrast and high resolution. Remember that pictures usually lose contrast during the printing process. Do not put together photographs and drawings in the same picture.
- If the pictures, drawings, illustrations or photographs included in the paper are copyrighted, it is responsibility of the author of the study to obtain the permission to use that material.
- Pictures, illustrations and photographs included in the paper must be gathered in one single folder named: "Figures_(the first letter of the first two words of the title of the paper in capital letters)." For example, if the title of the paper is "Population density in mainland Ecuador", the name of the folder should be: "**Figures_PD**".
- The general format for the identification labels of different material is:

For tables:
`\label{PD_tab:001}`

For figures:
`\label{PD_fig:001}`

For equations:
`\label{PD_eq:001}`

For definitions:
`\label{PD_def:001}`

For theorems:
`\label{PD_teo:001}`

etc.

Structure and Style

The structure of the paper must be as follows:

Mandatory	Optional
Title	Acknowledgements
Abstract	Recommendations
Keywords	Future studies
Resumen	Appendix
Palabras clave	
Classification	
Introduction	
Theoretical framework or Methodology	
Application or Results	
Conclusions	
References	

Tabla 2. Paper structure for evaluation.

Style of the document

For some parts of the document, various particularities must be considered.

Title. The Title must include the following information: name of the study, name of the author(s), addresses and emails of the author(s).

Abstract (in Spanish: Resumen). The Abstract must be in Spanish and English. It must be short and concise (250 words maximum); it must show the relevant results of the study, be informative and not indicative. For example:

“This paper analyzes the role played by individual, familiar and labor characteristics in the probability of obtaining a temporary or an indefinite contract,”

is informative.

“The probability of obtaining a temporary or an indefinite contract is discussed and accepted,”

is indicative and should be avoided.

Keywords (in Spanish: Palabras claves). There must be between 3 and 6 keywords that represent the main topics of the study. They must go at the end of the Abstract and the Resumen.

Classification. The paper must include the classification system from Journal of Economic Literature, **JEL**.

References. This section must include all the publications quoted in the text; it should not include unpublished articles or any other unverifiable information. Thesis should only be quoted if they are available in a physical or digital library. References must be in alphabetical order by last names. References to websites must follow a scheme similar to publications, including the access date. For example:

Referencias

- [1] M. AGUAYO, y E. LORE, *Cómo hacer una Regresión Logística binaria paso a paso II análisis multivariante*, Fundación Andalucía Beturia para la Investigación en Salud, Dot. N° 0702013, (2007).
- [2] C. GAMERO, *Satisfacción Laboral y tipo de contrato en España*, *Investigative Radiology* 34, Vol. 10, 636 - 642, (1999).
- [3] LEY No. 100, *Código de la niñez y adolescencia*, Registro Oficial 737, Enero, (2003).
- [4] SCIENTIFIC INSTRUMENT SERVICES, Inc. 2006. *MASS Spectral Library*. En línea: <http://www.sisweb.com/software/ms/nist.htm>, Consulta: 10 de abril del 2010.

Acknowledgements. They could be included as a whole section, at the end of the document, in a single paragraph.

Editorial procedures

Once the paper has been received by the Editor, it undergoes two qualifying processes. The first is performed by an internal editorial board which will determine the pertinence and scientific content of the paper. The second one consists of a *blind review* made by local and foreign experts. If the draft is accepted for publishing, it shall go through the following stages:

1. The Editor will inform the author of any observation made by the evaluators, so all the necessary corrections and changes would be made. This must not exceed a month.
2. Once the corrected paper has been received, the Editor will perform the copying-editing process. After the document has been formatted, a page proof will be sent to the author to check edition or format mistakes; the author shall not make any change to the document. The author must return the proof page, along with any observation, to the Editor one week after having receiving it.

3. The final draft will be kept until the printing process starts. The author will receive a Journal in hard copy and digital format.

If the paper is not accepted for publication, it will be returned to the author explaining the reasons for such decision.

Final Comments

The publication of the journal depends on the cooperation of authors, evaluators, printing shop and Editor; collaboration and meeting deadlines is fundamental.

Analítica Journal is published twice a year.

Código de ética

Es fundamental contar con el aval del Committee on Publication Ethics (COPE) lo que nos ha permitido mejorar las prácticas de ética de las publicaciones que se ha promovido en diferentes ámbitos de la investigación [4, 2], tanto para editores [3], autores [5] y lectores [1].

Editores

Responsabilidades generales de los editores:

- Mejorar continuamente la revista.
- Garantizar la calidad de los artículos publicados.
- Cuidar la integridad de su experiencia académica.
- Publicar correcciones, aclaraciones, retractaciones y disculpas de ser necesarias.
- Proteger la propiedad intelectual y derechos de autor.
- Mantener independencia editorial.

Responsabilidades de los editores con los lectores:

- Ofrecer un producto que considere los estándares internacionales manejados por las revistas científicas de mayor prestigio.
- Asegurar que el material publicado fue sometido a revisión y aprobación.

Responsabilidades de los editores con los autores:

- Consensuar la calidad del material a publicarse.
- Aceptar o rechazar un artículo para su publicación, basándose en su importancia, originalidad, claridad, relevancia para la revista.
- Informar sobre el proceso de revisión de sus artículos por parte de sus pares.
- Estar preparado para justificar los procesos de revisión, en caso de requerir un informe por parte del autor.
- Mantener actualizada la norma de publicación de trabajos.
- Cumplir la decisión de aceptación o rechazo de un artículo para su publicación.

- Editores nuevos no deben revocar las decisiones de publicar documentos presentados por el editor anterior, salvo excepciones (cuando va en contra de la imagen institucional y/o de la revista).

Responsabilidades con los revisores:

- Publicar y mantener actualizada la norma de revisión de trabajos.
- Mantener comunicación constante con los revisores.

Proceso de revisión por pares:

- Asegurar que el material remitido para su revisión es confidencial mientras lo examinan.

Autores

Responsabilidades generales de los autores:

- Estructurar el artículo cumpliendo el método científico, conforme reglas gramaticales y las normas de publicación de trabajo¹.
- Presentar suficientes referencias de fuentes públicas.

Responsabilidades de los autores con los lectores:

- Asumir la responsabilidad pública del contenido.
- Proporcionar a los lectores interesados copias de los datos, manual de procedimiento, material experimental.

Responsabilidades de los autores con los editores:

- Defender el contenido del artículo (datos, metodología, conclusiones).
- Mantener comunicación constante con el editor en jefe, para correcciones y revisiones.

¹Normas de presentación de trabajos, Analítica, Revista de Análisis Estadístico

Lectores

Responsabilidades generales de los lectores:

- Referenciar la bibliografía del artículo usado².
- Participar en el desarrollo y progreso de la revista, enviando sugerencias de los artículos publicados³.

Responsabilidades de los lectores con los editores⁴:

- Hacer pública discrepancias a través de cartas al editor.
- Señalar errores especificando el artículo y el autor.

Responsabilidades de los lectores con los autores⁵:

- Tener una actitud crítica con el contenido publicado señalando los errores.
- Enviar su opinión, crítica o discusión directamente a la dirección del autor/autores.

*Puede encontrar más recursos en COPE (Committee on Publication Ethics - Comité de Ética de Publicación) [1].

Referencias

[1] COMMITTEE ON PUBLICATION ETHICS, *Resources: Code of conduct*, COPE,

<http://publicationethics.org/resources/code-conduct>, (2011).

[2] INTERNATIONAL COMMITTEE OF MEDICAL JOURNAL EDITORS, *Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication*, ICMJE, www.icmje.org/urm_full.pdf, (2010).

[3] KLEINERT, S. Y WAGER, E., *Responsible research publication: international standards for editors*, 2nd World Conference on Research Integrity, World Scientific Publishing, (2010).

[4] SCOTT-LICHTER, D. Y EDITORIAL POLICY COMMITTEE, *Cse's white paper on promoting integrity in scientific journal publications*, Wheat Ridge, www.councilscienceeditors.org/files/public/entire_whitepaper.pdf, (2012).

[5] WAGER, E. Y KLEINERT, S., *Responsible research publication: international standards for authors*, 2nd World Conference on Research Integrity, World Scientific Publishing, (2010).

² A1. AUTOR1 y A2. AUTOR2, *Título del artículo*. Analítica, Vol N°, ISSN: 1390-6208, (año). páginas.

^{3 4 5} El lector debe estar debidamente identificado.

Analítica

volumen

Revista de Análisis Estadístico
Journal of Statistical Analysis

5

ISSN 1390-6208
e-ISSN 1390-7867
Número 1 - 2013
enero - junio

Administración Central

Quito
Juan Larrea N15-36 y José Riofrío
Teléfonos: (593) 22544 326 - 22544 561
Fax: (593) 22509 836 Casilla postal: 135C
correo-e: planta_central@inec.gob.ec
analitika@inec.gob.ec


instituto nacional de estadística y censos