

Comparabilidad y ajustes metodológicos de las estadísticas laborales: El caso de la tasa de desempleo en Ecuador*

Comparability and methodological adjustments of labor statistics: The case of the unemployment rate for Ecuador

JOSÉ GABRIEL CASTILLO**

CARLA MARÍA SALAS***

Resumen

Los institutos nacionales de estadística se ven en la necesidad permanente de reformar y actualizar los criterios e instrumentos estadísticos de análisis del mercado laboral, así como de adaptar las recomendaciones internacionales. Estos procesos plantean retos en cuanto a la comparabilidad de la información, necesaria para analizar dinámica del mercado laboral. El presente trabajo expone una aproximación metodológica consistente para el empalme de los agregados laborales a partir de información microeconómica de las encuestas de hogares. Analizamos el caso de la tasa nacional de desempleo para el Ecuador, evaluando las potenciales alternativas de simulación, y definimos un criterio de selección consistente para el pronóstico y empalme de la serie.

Palabras clave: Desempleo, empalme de indicadores laborales, mercado laboral, encuestas de empleo.

Clasificación JEL: C25, J21, J24, J64.

Abstract

The national statistical institutes face permanently the need to reform and update the statistical criteria and instruments for labor market analysis, as well

* Agradecemos los comentarios de Roberto Castillo, Ana Rivadeneira y los técnicos de la Dirección de Estudios y Análisis de la Información del INEC; así como la gentil colaboración de Mariuxi Arias del SIISE, cuya contribución esclareció la hoja de ruta de este trabajo. Agradecemos también los comentarios del editor y un revisor anónimo, que contribuyeron a mejorar sustancialmente este documento. Los errores son responsabilidad exclusiva de los autores.

** Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Sociales y Humanísticas, Guayaquil. E-mail: jcastil@espol.edu.ec (Autor correspondiente).

*** Investigadora independiente. E-mail: cmsalasc@gmail.com

as to adapt international recommendations for measurement. These processes pose important challenges regarding the comparability of information, necessary to analyze the labor market dynamics. This study presents a consistent methodological approach for the splicing of labor market aggregates based on microeconomic information from household surveys. We analyze the case of the national unemployment rate for Ecuador; assess the potential alternatives for simulation and define a consistent selection criterion for the prognosis and splicing of the series.

Key words: *Unemployment, splicing of labor indicators, labor market, labor surveys.*

JEL Classification: *C25, J21, J24, J64.*

1. INTRODUCCIÓN

Uno de los requerimientos fundamentales de cualquier análisis dinámico de los agregados laborales es la comparabilidad de los indicadores. Con frecuencia, los institutos nacionales de estadística se ven en la necesidad de reformar sus instrumentos estadísticos, procedimientos de recolección de información, metodologías de cálculo de indicadores, etc., con el objetivo de mejorar la capacidad de descripción y análisis del estado y evolución del mercado laboral. Asimismo, múltiples instancias como la Conferencia Internacional de Estadísticos del Trabajo (CIET), en el marco de la Organización Internacional del Trabajo (OIT), establecen periódicamente recomendaciones y directrices que buscan armonizar los mecanismos de medición. Mediante la implementación de mejores prácticas acerca de los criterios de evaluación e indicadores generados, se busca que la información sea consistente y comparable entre países. Los cambios propuestos no están exentos de críticas tanto de carácter técnico como político-institucional, sobre todo cuando estos impiden la comparabilidad de la información y los quiebres observados no se relacionan con la dinámica real del mercado. No en pocas ocasiones los investigadores o las instituciones multilaterales, que analizan información consolidada de varios países, tienen que adaptar mecanismos *ad hoc* para superar las limitaciones de comparabilidad de las cifras; tarea que puede ser directa –si los instrumentos estadísticos lo permiten– o muy compleja, en cuyo caso el “criterio de experto” es la única guía¹.

Este es el caso de Ecuador. En el 2007 el Instituto Nacional de Estadísticas y Censos (INEC) del país implementó una nueva metodología para la construcción de los indicadores del mercado laboral, con dos objetivos fundamentales: i) unificar las estadísticas laborales derivadas del levantamiento de información por parte de un ente externo; y, ii) acogerse a los estándares internacionales sugeridos por la CIET. Este nuevo marco conceptual presenta algunos cambios en los conceptos y preguntas del instrumento de recolección de la información (formulario) en la Encuesta Nacional de Empleo y Desempleo (ENEMDU), principal fuente de información de las estadísticas laborales en el país. Este

¹ Ver por ejemplo Ball *et al.* (2013).

cambio metodológico, favorable en términos de precisión, comparabilidad internacional y claridad de interpretación de los indicadores laborales, trajo consigo una consecuencia perniciosa: la imposibilidad de analizar la dinámica del mercado laboral por la falta de comparabilidad de las cifras².

El cambio metodológico produjo amplia inconformidad en analistas económicos, tomadores de decisión y usuarios generales de la información de los agregados laborales. En parte, esto responde al desconocimiento de las recomendaciones y tendencia de la normativa internacional, pero, sobre todo, a la imposibilidad de contrastar la real evolución de los indicadores del mercado laboral en el período pre y poscambio³.

Las instituciones nacionales vinculadas al levantamiento, procesamiento y análisis de información; como el Banco Central del Ecuador (BCE) y en particular el INEC, han realizado algunas propuestas metodológicas que consideran esta necesidad; no obstante, por diversas razones no existen cifras oficiales del mercado laboral que sean comparables entre las metodologías pre y post 2007⁴.

El presente documento expone un ejercicio de consolidación de la tasa nacional de desempleo del Ecuador, comparable para el período 2000-2016. A partir del uso de información microeconómica de las encuestas de hogares, la metodología propuesta se basa en la reclasificación de los individuos en las categorías de desempleados e inactivos, respondiendo fundamentalmente al cambio en el período de referencia de 5 a 4 semanas de búsqueda de empleo. La propuesta presentada considera tanto un criterio óptimo de selección del modelo a emplear como el empalme de las series desde dos perspectivas: la aplicación de la metodología nueva al período previo, y la extensión de los criterios de la metodología antigua a las estadísticas poscambio. La conclusión final, a base de lo observado en la dinámica de la serie empalmada, es que los niveles de desempleo actuales coinciden con los niveles observados en los años 2004-2006, el período de recuperación de la crisis financiera, planteando retos pendientes en la capacidad de generación de empleo de la economía en el período de análisis.

Lo que resta del documento está organizado de la siguiente forma. La segunda sección describe brevemente las experiencias previas, fundamentalmente el caso colombiano, referencia de nuestro análisis. La tercera sección resume los principales cambios metodológicos del caso ecuatoriano. La cuarta sección describe la metodología de empalme de las series de desempleo, las fortalezas y debilidades de la aproximación propuesta. En la quinta sección se presentan y analizan los resultados, y la sexta sección concluye el documento.

² En caso de la ENEMDU, por ejemplo, otros cambios importantes posteriores a los de diciembre del 2017 incluyen: migración a un nuevo marco de muestreo en el 2013, incremento del tamaño de la muestra en el 2014, entre otros (INEC).

³ Este sentimiento se vio agravado por la coincidencia del cambio metodológico con el cambio de gobierno, lo que impide evaluar adecuadamente la gestión gubernamental en términos de la generación o destrucción de empleo.

⁴ Hasta donde conocemos, el ente rector se encuentra trabajando en el empalme a base de un ejercicio que considera criterios de experiencias internacionales. Algunas de estas experiencias se recogen en este documento; sin embargo, al momento de elaboración de este trabajo no se han publicado resultados oficiales.

2. EXPERIENCIAS INTERNACIONALES PREVIAS

Existen múltiples experiencias internacionales que engloban cambios metodológicos que devienen en modificaciones de los agregados laborales. Guataquí y Tabora (2006) exponen una breve reseña de los cambios metodológicos en el caso británico, español, italiano y de Trinidad y Tobago. La mayoría de estas modificaciones responden a las recomendaciones de la OIT; no obstante, es importante recordar que los países ejercen soberanía estadística en cuando a la adaptación de las recomendaciones y estándares internacionales, lo que conlleva a que la estandarización no esté exenta de críticas o problemas de carácter técnico, además de generar imprecisión en la comparabilidad de las cifras.

En la región, varios países han emprendido estos cambios metodológicos en la última década. El caso más cercano al Ecuador es el caso colombiano. En Colombia, en el 2000, por medio del Departamento Administrativo Nacional de Estadística (DANE) se realizaron cambios metodológicos profundos en busca de acoger las recomendaciones de las organizaciones internacionales involucradas en las estadísticas laborales, principalmente la OIT. La Encuesta Nacional de Hogares (ENH), vigente desde 1976 y levantada de forma trimestral, similar al caso ecuatoriano, fue reemplazada en el 2001 por la Encuesta Continua de Hogares (ECH), con una periodicidad mensual mediante el Sistema Continuo de Recolección de Datos. Adicionalmente se realizaron cambios en la cobertura de la muestra y definiciones que afectaron a los agregados laborales, en particular a la categoría de desocupados, mediante la modificación de la Población en Edad de Trabajar (PET). Este cambio metodológico significó una reducción de aproximadamente 2,8 puntos porcentuales en la tasa de desempleo (Arango 2006).

Los cambios más importantes observados en el caso de Colombia son:

- Cambios en la definición de los Trabajadores Familiares Sin Remuneración (TFSR). A partir del 2001, si un individuo trabajaba al menos 1 hora a la semana en negocios familiares sin remuneración pasa a formar parte de la población ocupada, mientras que en la metodología previa se clasificaban como TFSR a quienes destinaban al menos 15 horas semanales en dichas condiciones.
- Cambios en la clasificación del desempleo abierto y el desempleo oculto. A partir de la ECH se incorporan dos cambios relevantes para la clasificación: i) el criterio de disponibilidad inmediata para trabajar, y; ii) el período de referencia para la búsqueda que pasa de 1 semana a 4 semanas.
- Cambios en el proceso operativo de recolección de datos e implementación de cartografía para el trabajo de campo.

Como resultado de estas modificaciones se evidenciaron cambios significativos en la evolución de los indicadores laborales colombianos. En búsqueda de armonizar las metodologías, Lasso (2002) aplica modelos de probabilidad para la reclasificación de los TFSR de acuerdo con el número de horas trabajadas a la semana (menos de 15 horas), así como el criterio de disponibilidad inmediata para la posterior homogenización de las series temporales. Su aproximación corrige la influencia del factor estacional, a partir de las encuestas de hogares y las definiciones de población ocupada, desempleada e inactiva. Arango *et al.* (2006) expanden este análisis remediando potenciales sesgos de variables omitidas

en los factores determinantes de la reclasificación; por ejemplo, incluyendo en las estimaciones otros controles relacionados a: ciudad, grupos de edad, entre otros; así como múltiples factores estacionales. Adicionalmente, los autores utilizan una segunda metodología de empalme a partir de datos agregados del mercado laboral (aproximación macro). Observan que aproximadamente 11% de reclasificación de los desocupados como inactivos entre el 2001 y 2005 se deriva del cambio en el criterio de disponibilidad inmediata, lo que se traduce en una reducción de la tasa de desempleo de aproximadamente 1,9 puntos porcentuales; mientras que solamente 0,1 puntos porcentuales devienen de los cambios en los TFSR. En cuanto a la metodología empleada los autores concluyen que “...el modelo microeconómico (*logit*) presenta un mejor desempeño como lo evidencia el ejercicio realizado para las siete principales ciudades; más aún, un análisis estadístico de los momentos muestrales de cada serie y su poder de pronóstico sugiere que el modelo microeconómico (*logit*) es el más adecuado para el empalme de las series del mercado laboral”.

3. EL CASO ECUATORIANO

En el Ecuador, en junio del 2007, el Banco Central del Ecuador, que hasta entonces coordinó la Encuesta de Coyuntura del Mercado Ecuatoriano⁵, firma un convenio con el Instituto Nacional de Estadísticas y Censos (INEC) para que sea esta institución la encargada del levantamiento de información. El objetivo final fue el de homologar los distintos instrumentos estadísticos del mercado laboral ecuatoriano. Paralelamente, los equipos técnicos de ambas instituciones trabajaron en la implementación de una nueva metodología para la construcción de los indicadores del mercado laboral ecuatoriano, con dos objetivos fundamentales; por un lado, unificar las estadísticas laborales derivadas del levantamiento de información por parte de un ente externo y, por otro, acoger los estándares internacionales sugeridos por la Conferencia Internacional de Estadísticos del Trabajo (CIET), en el marco de la OIT⁶.

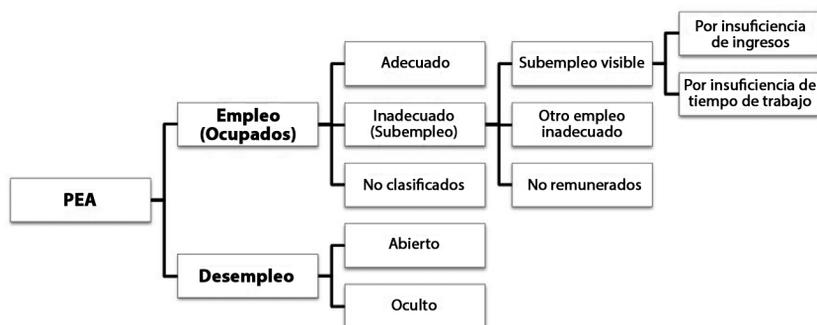
A partir de septiembre del 2007, el proceso operativo de levantamiento de información pasó a manos del INEC, ente rector de las estadísticas nacionales. Al mismo tiempo, se implementa el Sistema Integrado de Encuestas de Hogares (SIEH) cuyos dos cambios más relevantes son: la modificación del marco muestral a partir del VI Censo de Población y V de Vivienda del año 2010, en donde la representatividad muestral se amplía a nivel provincial, regional, urbano-rural y cinco ciudades principales: Quito, Guayaquil, Cuenca, Machala y Ambato (se excluye la región insular). También se modifica la definición de urbano y rural; a partir de este momento se considera como urbana cualquier población con más de dos mil habitantes.

⁵ Encuesta levantada conjuntamente con la Facultad Latinoamericana de Ciencias Sociales-FLACSO, sede Quito.

⁶ Las CIET son conferencias celebradas aproximadamente cada cinco años en el marco de la OIT para definir estándares y plantear recomendaciones y normas internacionales respecto de las estadísticas laborales, con el fin de facilitar la comparabilidad de los indicadores en los países miembros. La adaptación del marco conceptual se deriva de la 13va y 16va CIET (INEC, 2014).

Además de modificar el marco muestral y los procedimientos de levantamiento de encuestas en los operativos de campo, los cambios son fundamentalmente conceptuales. Se modifican las definiciones de las preguntas de los formularios de la encuesta con los que se construyen los indicadores laborales. Estos cambios afectan a la conformación de la Población Económicamente Activa (PEA) y por tanto a la Población Económicamente Inactiva (PEI). La Figura 1 describe las categorías de la nueva metodología y la Tabla 1 resume los cambios conceptuales más relevantes que influyen en la construcción de los agregados laborales.

FIGURA 1
CATEGORÍAS DE CLASIFICACIÓN DE LA PEA: NUEVA METODOLOGÍA



Fuente: INEC (2014), BCE (2008).
Elaboración: Autores.

TABLA 1
RESUMEN DE CAMBIOS METODOLÓGICOS RELEVANTES

Categoría	Antigua metodología	Nueva metodología
Ocupados	Plenos y subempleados	Se incorpora adicionalmente "Ocupados no clasificados"
Subempleo	Visible, invisible e informal	Visible, otro empleo inadecuado, y no remunerado
Subempleo visible	Trabajan menos de 40 horas y no pertenecen al sector informal	Cumplen con tres criterios: 1. Desea trabajar más 2. Está disponible para trabajar más horas; y, 3. Debe haber trabajado menos de 40 horas.
Subempleo invisible	Personas que trabajan 40 horas o más pero tienen ingresos menores al salario legal unificado y no pertenecen al sector informal	Se elimina

Categoría	Antigua metodología	Nueva metodología
Subempleo informal	Todos los ocupados del sector informal	Se elimina
Otras formas de subempleo	No existe	Personas ocupadas que: trabajan 40 horas o más con ingresos superiores al salario unificado legal y que están disponibles; trabajan 40 horas o más con ingresos menores al salario unificado legal; trabajan menos de 40 horas con ingresos menores al salario unificado legal y no están disponibles.
Desempleo abierto	Personas sin empleo y que en la semana de referencia buscaron activamente empleo en las 5 SEMANAS ANTERIORES A LA FECHA DE LA ENTREVISTA.	Personas sin empleo y que en la semana de referencia buscaron activamente empleo en las 4 SEMANAS ANTERIORES A LA FECHA DE LA ENTREVISTA.
Desempleo oculto (Distintas razones)	Personas sin empleo y que en la semana de referencia NO buscaron activamente empleo en las 5 SEMANAS ANTERIORES A LA FECHA DE LA ENTREVISTA. 7 razones de no búsqueda iniciales.	Personas sin empleo y que en la semana de referencia NO buscaron activamente empleo en las 4 SEMANAS ANTERIORES A LA FECHA DE LA ENTREVISTA. Se modifica la redacción de las 7 razones de no búsqueda.
Inactivos	Distintas razones de clasificación	Todas las personas de 15 años y más, no clasificadas como ocupadas o desempleadas durante la semana de referencia y que incluyen: rentistas, jubilados, pensionistas, estudiantes, amas de casa, entre otros.

Fuente: INEC (2014), BCE (2008).

Elaboración: Autores.

Uno de estos cambios, en el que no existe un mecanismo instrumental adecuado que permita una eficaz reclasificación de los individuos, es precisamente el relacionado al desempleo (abierto y oculto). Como muestra la Tabla 1, el período referencial de búsqueda de empleo, para la clasificación de un individuo como desempleado o inactivo, cambia entre metodologías. Mientras en la metodología “antigua” el período de referencia para la búsqueda de empleo fue de 5 semanas anteriores a la fecha de la entrevista, este período se redujo en una semana en la “nueva” metodología, pasando a ser de 4 semanas (ver más detalles en sección empírica). La consecuencia obvia de este cambio es una reducción del número de individuos en la categoría de desempleados. Por ejemplo, un in-

individuo entrevistado pudo buscar activamente empleo en las últimas 4 semanas sin conseguirlo, en cuyo caso, en ambas metodologías sería clasificado como desempleado. Por el contrario, si este mismo individuo NO buscó trabajo en las últimas 4 semanas (por razones identificadas), pero sí lo hizo una semana antes de las 4 de referencia (semana 5), en la “antigua” metodología sería clasificado como desempleado mientras que en la “nueva” pasaría simplemente a formar parte de la población inactiva (PEI).

Como resultado del cambio metodológico, la tasa de desempleo cayó en un solo año en aproximadamente 1,3 puntos porcentuales. Esto generó un quiebre en la serie temporal pues la tendencia a partir del cambio metodológico mantiene niveles promedio muy por debajo del período previo; aproximadamente 2,6 puntos porcentuales de diferencia entre el promedio del 2000 al 2006 y del 2007 al 2016.

4. METODOLOGÍA DE EMPALME

La metodología de cálculo de los agregados laborales en el Ecuador a partir del 2007, en particular de la tasa nacional de desempleo, modifica fundamentalmente dos aspectos: i) el período de referencia para la reclasificación, esto es, el período de búsqueda de empleo que se reduce a 4 semanas en total a partir de la semana completa antes de la fecha de la entrevista (antes 1 + 4 semanas)⁷, y; ii) la clasificación de empleados o inactivos, según la disponibilidad y deseo de trabajar.

El primer aspecto constituye el punto central en el que se enfoca esta investigación. Para identificar la población desempleada, la ENEMDU hasta el año 2006 incluyó preguntas como: “¿buscó trabajo la semana pasada?”, “¿estuvo buscando trabajo **las cuatro semanas anteriores a la semana pasada**?”, “¿hace cuánto tiempo busca trabajo?”, “¿por qué motivos no buscó trabajo?” y “¿desea trabajar y está disponible para hacerlo?”. En su lugar, a partir del 2007 se incluyen las siguientes preguntas: “¿en las **últimas cuatro semanas** ha realizado alguna gestión para aumentar las horas de trabajo o cambiar de trabajo?”, “¿hace cuánto tiempo busca trabajo?”, “¿por qué razón no buscó trabajo?” y “¿está disponible para trabajar?”. La pregunta central que dificulta la identificación de la población desempleada es la que hace referencia al período de búsqueda de empleo que cambia de cinco semanas a cuatro. Las preguntas de identificación para la población de empleados no se modifican con la nueva metodología. Por tanto, es necesario establecer una estrategia que permita identificar a los individuos según su decisión de búsqueda en esa semana adicional de la que no se dispone de información. El ejercicio de reclasificación para la homogenización de las series de desempleo requiere de un marco que explique la decisión de participación en el mercado laboral de los individuos. De esta forma, motivamos la aproximación empírica a partir de un enfoque neoclásico de participación en la fuerza laboral (oferta laboral).

⁷ Recordemos que se trata de una encuesta y por tanto la información es sensible a la capacidad del entrevistado de recordar las gestiones realizadas en el período de referencia. En principio, al ser menor el período, mayor capacidad de recordación.

4.1. Intuición teórica

El modelo canónico de oferta laboral plantea la dicotomía de la elección entre las horas de trabajo necesarias para mantener un determinado nivel de consumo (C) y la utilidad generada por el disfrute de las horas de ocio (I). El nivel de satisfacción obtenido a base de las decisiones de asignación del tiempo se resume en una función de utilidad simplificada: $U(C, I)$. Estas decisiones de asignación de tiempo, y por tanto, de participación en el mercado laboral, dependen de si la utilidad del nivel consumo que facilita el salario de mercado –el costo de oportunidad del ocio– es mayor a la utilidad del nivel de consumo posible mediante el salario de reserva –el incremento mínimo de ingreso para trabajar la primera hora–; condicional en otros factores, como ingresos no laborales, y tomando en consideración las preferencias, representadas en la forma funcional de la utilidad. El supuesto fundamental de comportamiento es que los individuos determinan esa asignación de tiempo en función de maximizar su bienestar.

El salario de reserva de un individuo constituye una variable latente (no observada) y, por tanto, la información disponible no permite determinar si este es mayor, menor o igual al salario de mercado (observado). De esta forma, la diferencia entre el salario de mercado (w^m) y el de reserva (w^r) es también una variable latente que denotaremos por y^* :

$$(1) \quad y^* = w^m - w^r$$

El individuo busca trabajo únicamente si $y^* > 0$, es decir, si el salario de mercado es mayor al salario de reserva; y deja de buscarlo si $y^* \leq 0$. La teoría sugiere que, frente a un salario fijado por el mercado, a mayor salario de reserva menor es la probabilidad de que un individuo busque trabajo; esto debido a que el bienestar que le genera el salario de mercado no justifica modificar su condición laboral y obtener empleo. Asimismo, condicional en un nivel de salario de reserva establecido (w_0^r), incrementos en el nivel de salario de mercado incrementan la probabilidad de que un individuo participe del mercado laboral (Borjas, 2016).

Para identificar la condición laboral de un individuo es necesario distinguir qué factores están correlacionados en la caída del salario de mercado, provocando que individuos considerados antes como desempleados en la antigua metodología (periodo de tratamiento) sean reclasificados como inactivos en la nueva, ya que no buscaron trabajo en el nuevo período de referencia establecido en la encuesta. Si $y^* \leq 0$, el individuo se clasifica como inactivo y, por tanto, podemos modelar la decisión de búsqueda de trabajo evaluando la probabilidad de participar del mercado laboral a partir de una variable dicotómica observada y , tal que:

$$(2) \quad P(y = 1) = P(y^* \leq 0)$$

en donde y toma el valor de 1 si el individuo es inactivo, es decir, no buscó trabajo en el período de análisis por razones identificadas en la encuesta. Como corresponde, la categoría alternativa, cuando $y=0$, o individuo activo, es mutuamente excluyente y representa a los individuos que al decidir buscar empleo no lo encuentran, esto es, los desempleados.

La realización de la variable latente puede modelarse en función de las características del individuo y los factores que determinan la oferta laboral. Por tanto tenemos que:

$$(3) \quad y^* = X_i' \gamma + \epsilon_i$$

en donde el vector X_i resume toda la información disponible en las encuestas para el individuo i ; el vector γ representa los parámetros de estimación y el término ϵ_i el error idiosincrático. Se evalúa la probabilidad de que un individuo i no busque trabajo (sea inactivo), a partir de la distribución de los errores. Así tenemos que:

$$(4) \quad P(y = 1) = P(y^* \leq 0) = P(X_i' \gamma + \epsilon_i \leq 0) = P(\epsilon_i \leq -X_i' \gamma) = F(-X_i' \gamma)$$

La selección de la densidad acumulada $F(\cdot)$ tiene varias alternativas. Para la aplicación empírica preferimos una distribución logística (modelo *logit*) por tradición (relacionada al Modelo de Utilidad Aleatoria) y facilidad de exposición e interpretación de los coeficientes estimados; no obstante, se puede reemplazar esta especificación asumiendo distribuciones alternativas; por ejemplo, la distribución normal (modelo *probit*). Ambas especificaciones de reclasificación arrojan resultados similares y las conclusiones de nuestro estudio no se alteran. De esta forma, para cada individuo la distribución empleada se ve de la siguiente forma:

$$(5) \quad P(y_i = 1 | X_i' \gamma) = \frac{1}{1 + e^{-X_i' \gamma}}$$

Asimismo, una especificación conveniente para la directa interpretación de los parámetros estimados es el logaritmo de la razón de probabilidades (*log odds ratio*, por sus siglas en inglés), que es:

$$(6) \quad \ln \left(\frac{P(y_i = 1 | X_i' \gamma)}{P(y_i = 0 | X_i' \gamma)} \right) = X_i' \gamma$$

en donde, la razón de probabilidades, entre ser inactivo o desempleado, varía en función de una combinación lineal de los factores que intervienen en la decisión de búsqueda de empleo. Los coeficientes estimados (γ) representan el cambio marginal en el *ratio* de probabilidades entre ser inactivo o desempleado.

Una aproximación análoga, que se concentra en modelar el tiempo de búsqueda ($T > 0$), en lugar de la decisión acerca del salario de reserva, es el uso de un modelo de duración o análisis de supervivencia. En este planteamiento el suceso de análisis es el período de búsqueda de empleo. Empleando las variables explicativas se puede modelar los parámetros de la probabilidad de que un individuo de la muestra se mantenga activamente en la búsqueda de trabajo en un período de referencia ($P(t \leq T \leq t+h | T \geq t)$), según la metodología de análisis: 4 o 5 semanas, para la nueva y vieja metodología, respectivamente. Por ejemplo, condicional en haber buscado empleo en las últimas 4 semanas (t), ¿cuál es la

probabilidad de que continúe en la búsqueda una semana adicional (h)?. De esta forma, se puede modelar si un individuo se clasifica como activo o inactivo, en función del tiempo que dura la transición de búsqueda de empleo (*hazard rate*). Condicional en disponer de información específica respecto del tiempo de búsqueda de empleo (T), esta aproximación es potencialmente preferible, porque modela directamente una variable aleatoria cuya distribución se puede asumir de manera paramétrica (ej. exponencial, Weibull, log-normal, etc.) o no paramétrica. No obstante, debido a la información disponible en las encuestas de empleo, su aplicación para el empalme en el caso de Ecuador es limitada. En particular, la encuesta no recoge el tiempo de búsqueda de empleo de los individuos, variable necesaria para una modelización consistente en el análisis de duración de empleo⁸. Por el contrario, la encuesta recoge una respuesta binaria a una pregunta cuyo tiempo de referencia está preestablecido por la metodología de levantamiento de información (ver introducción de la sección 4). Esto reduce el problema de análisis a un modelo binario, por lo que favorecemos la aproximación teórica y empírica propuesta, sin adicionar mayor estructura que la que brinda la disponibilidad de información.

4.2. Metodología empírica

El objetivo fundamental de la aplicación es corregir el nivel de desocupados y , por consiguiente, el nivel de la Población Económicamente Inactiva-PEI en los agregados laborales. Partiendo de los argumentos expuestos, utilizamos un modelo de probabilidad *logit* de reclasificación entre desempleados e inactivos. Esta aproximación busca homologar el cálculo de la tasa nacional de desempleo para ambos períodos (pre y post-2007) a partir del pronóstico de la probabilidad de que un individuo desempleado sea reclasificado como inactivo, en el período de tratamiento (el período de aplicación de la simulación o predicción fuera de la muestra), a partir de información observable de las características de los agentes inactivos, en el período de control (el período de referencia para la estimación de los coeficientes).

A partir del planteamiento teórico, el ejercicio de estimación se basa en la maximización de la función de verosimilitud $L(\gamma)$ estándar para un modelo *logit*, a partir de la ecuación 5; así tenemos⁹:

$$(7) \quad \max_{\gamma} L(\gamma) = \prod_{i=1}^N P(y_i = 1 | X_i' \gamma)^{y_i} \times \left[1 - P(y_i = 1 | X_i' \gamma) \right]^{(1-y_i)}$$

en donde y_i representa la variable dicotómica de si el individuo es desempleado ($y_i=0$) o inactivo ($y_i=1$). El vector de covariantes X_i contiene múltiples combi-

⁸ También es posible modelar la función de riesgo a partir de convertir en intervalos discretos una variable continua. Sin embargo, nuevamente la falta de variabilidad en la variable de respuesta impide aprovechar esta aproximación. Agradecemos a un revisor anónimo por sugerir este enfoque.

⁹ Alternativamente, se maximiza la transformación monótonica a partir del logaritmo de la función de verosimilitud ($\ln(L)$).

naciones de variables que incluyen: ciudad de residencia, quintiles de ingresos, estrato socioeconómico, número de ocupados en el hogar, presencia de niños menores a 6 años, estado civil, años de escolaridad, asistencia escolar, edad, género, escolaridad del jefe del hogar, parentesco con el jefe del hogar, área de residencia (urbano/rural) y variables geográficas (provincia). Adicionalmente, algunos modelos incluyen interacciones entre años de escolaridad y las variables: ciudad de residencia, género y área (modelo saturado). La variable ciudad de residencia está conformada por cuatro variables *dummy* en donde se destacan las ciudades principales: Guayaquil, Quito, Cuenca y Otras. Los quintiles de ingreso se obtienen a partir del ingreso per cápita del hogar¹⁰. El nivel socioeconómico se clasificó en cinco niveles tomando como referencia el cuestionario de estratificación de nivel socioeconómico desarrollado por el INEC (2011). Este cuestionario consiste en un sistema de puntuación en donde las características de la vivienda tienen un puntaje máximo de 216 puntos, educación 171 puntos, economía 170 puntos, bienes 163 puntos y tecnología 80 puntos. Cada nivel socioeconómico está representado por una variable *dummy*, excepto la categoría base. En cuanto a la información de estado civil de los individuos, por inconsistencias de registro entre bases de datos se optó por una aproximación en donde se agrupan los individuos que tienen pareja o cónyuge, versus los que no la tienen. Para las variables de parentesco con el jefe del hogar se realizó un ejercicio de filtración y consolidación tomando en cuenta los miembros del hogar, y criterios razonables de identificación según las preguntas de la encuesta.

Es importante tomar en cuenta los efectos estacionales y la dinámica de los indicadores de empleo dentro de la estimación, de lo contrario se arriesga obviar efectos dinámicos en el mercado laboral que pueden afectar la reclasificación. Las especificaciones empleadas incluyen la tendencia anual, o efectos fijos por año de análisis, controlando de esta forma los efectos contemporáneos comunes de los indicadores laborales. No obstante, también es posible la presencia de cambios estructurales en los parámetros del modelo de pronóstico, si existen efectos conjuntos entre regresores variables en el tiempo. Recordemos que, *a priori*, el único cambio relevante es el cambio metodológico objeto de este estudio. Precisamente la dificultad de comparar las cifras impide evaluar la presencia de otros factores estructurales en la dinámica laboral¹¹. Desafortunadamente, todos los sistemas de información estadística del país presentan información únicamente a partir del 2007, lo que dificulta el análisis¹². Para evitar asumir mayor estructura en el

¹⁰ La muestra se divide en cinco partes iguales, en donde el quintil de ingreso *q1* representa la porción de la muestra más pobre y el quintil de ingreso *q5* la porción de la muestra con mayores recursos. Se incluye una variable *dummy* por cada quintil de ingreso excepto la categoría de referencia, el quintil 1. Preferimos analizar la distribución de los quintiles en lugar del ingreso, ya que al tratarse de una encuesta, la variable de ingresos está típicamente sujeta a errores de medición y revelación; sin embargo, la asignación a un quintil de ingresos es menos sensible a este problema, mientras conserva la capacidad explicativa del mismo fenómeno. Si en su lugar se emplea directamente el ingreso, el análisis y resultados no se alteran.

¹¹ Asimismo, aun considerando los cambios metodológicos, no hay suficiente evidencia estadística de un quiebre estructural de la serie, aunque la tendencia se modifica. Ver sección 5.3.

¹² Ver por ejemplo SIISE: <http://www.siise.gob.ec/siiseweb/>

pronóstico, procedemos a emplear la información microeconómica disponible en las encuestas, tomando en cuenta múltiples especificaciones.

Asimismo, es necesario recalcar que la estimación del modelo de selección propuesto presupone la exogeneidad de los regresores relevantes; en otras palabras, el planteamiento parte del supuesto de que no existe ningún fenómeno o variable ausente en la especificación que afecte los regresores de la ecuación (ausencia de variables omitidas); que los regresores representan de manera fidedigna el fenómeno a caracterizar (ausencia de errores de medición), y que la decisión de buscar trabajo no determina los regresores incluidos (no simultaneidad o causalidad inversa). En síntesis, presupone la exogeneidad estricta de los errores. Bajo estos supuestos, la estimación por Máxima Verosimilitud (ecuación 7) es consistente y los estimadores son eficientes. No obstante, partir de estos supuestos no blindará la estimación de la presencia de potenciales sesgos. Al contrario, conscientes de esta limitación, compartida en cualquier aproximación empírica que carezca de un diseño de investigación (*research design*) adecuado para el análisis, procedemos a emplear el modelo propuesto enfatizando su capacidad predictiva por sobre su análisis causal. Dicha capacidad es evidente en la precisión de la reclasificación de los individuos en las categorías de interés, de acuerdo con los puntos de corte óptimos dentro de la muestra de referencia. El objetivo de estudio dista mucho de estudiar una relación causal entre los regresores incluidos en el modelo y la decisión de buscar empleo (los determinantes de la oferta laboral); al contrario, busca proponer una alternativa consistente de reclasificación de los individuos para el cálculo de las tasas de desempleo y empalme de las series afectadas por cambios metodológicos relevantes. La metodología propuesta, a nuestro criterio, cumple con el objetivo.

Otro aspecto relevante de la metodología se relaciona con la selección del período de referencia para la aplicación del proceso de homologación de la tasa nacional de desempleo. Dos aproximaciones metodológicas son posibles para el empalme de la serie:

1. *Reclasificación hacia atrás*: A partir de la nueva metodología, se puede estimar un modelo de reclasificación cuyos parámetros se estiman en la muestra de control y corresponden a los años 2007-2016. El pronóstico se realiza en los años de tratamiento, período 2000-2006.
2. *Reclasificación hacia adelante*: A partir de la antigua metodología, el modelo de reclasificación se estima en la muestra de control que corresponde a los años 2000-2006. El pronóstico se realiza en el período 2007-2016, que en este caso pasa a ser la muestra de tratamiento.

Ambas aproximaciones tienen ventajas y desventajas puntuales. La reclasificación hacia atrás tiene la ventaja de constituir un ejercicio directo y es preferible desde varios puntos de vista. Por un lado, el proceso de reclasificación está fundamentado principalmente en el desconocimiento de la decisión de búsqueda en la semana adicional a las 4 semanas a partir de la fecha de la entrevista (5ta semana). La metodología nueva se concentra en obtener información únicamente respecto de las 4 semanas y, por lo tanto, la metodología antigua absorbe, en el sentido temporal, la decisión de búsqueda de la nueva metodología. Por otro lado, para la evaluación de las políticas económicas en tiempo real, así como la

comparabilidad respecto de las mediciones internacionales vigentes, la reclasificación hacia atrás ofrece un ejercicio más consistente, útil, y por tanto preferible.

La reclasificación hacia adelante ofrece un enfoque distinto desde el punto de vista intuitivo y comunicacional. Este enfoque atiende la demanda de los tomadores de decisiones y analistas al ofrecer una compatibilización de los indicadores de empleo, independiente de cualquier agenda implícita que los cambios metodológicos hayan motivado (ej. agenda política). Contrario a la aproximación previa, la segunda aproximación es más compleja en la implementación, ya que involucra mayor incertidumbre respecto del criterio de reclasificación; la metodología *nueva* no contiene información respecto de la semana adicional.

Conscientes de la imposibilidad –justificada– de realizar ejercicios paralelos de levantamiento de información por parte del ente rector de las estadísticas nacionales¹³, realizamos ambos ejercicios y analizamos los resultados concentrándonos en la dinámica de la tasa nacional de desempleo.

Finalmente, el uso del modelo para el pronóstico requiere establecer criterios razonables de selección. Como menciona Arango *et al.* (2006), una de las limitaciones de esta aproximación es que no permite una reclasificación única. La simulación es sensible tanto a la especificación del modelo a emplear como a la decisión que se tome respecto del punto de corte de la probabilidad para la asignación del individuo a una u otra categoría (desempleado o inactivo). Para la definición de un criterio de optimalidad que guíe la selección del modelo a utilizar proponemos un procedimiento que tome en cuenta un criterio de minimización del error de predicción de la tasa nacional de desempleo en la muestra de control; es decir, que minimice el Error Cuadrático Medio de la Predicción (*Mean Square Error-MSE*, por sus siglas en inglés), de la siguiente forma¹⁴:

$$(8) \quad MSE = T^{-1} \sum_{j=1}^T (\hat{\tau}_j - \tau_j)^2$$

en donde, τ_j representa la tasa nacional de desempleo observada (calculada por el INEC) en el año j ; $\hat{\tau}_j$ representa la tasa nacional de desempleo predicha a partir de la reclasificación de individuos y basada en el pronóstico de la probabilidad individual de pertenecer al grupo de inactivos; y T representa el año final de la muestra de control. Por tanto, el intervalo de la sumatoria cambia de $j=2007$ hasta $T=2016$, para la reclasificación hacia atrás, y; $j=2000$ hasta $T=2006$, para la reclasificación hacia adelante.

¹³ Técnicamente no es posible levantar dos formularios con las mismas unidades de investigación (hogares), porque esto derivaría en sesgos de recolección relacionados al primer formulario aplicado, anulando la veracidad de la información. Asimismo, levantar formularios múltiples con muestras distintas constituiría un costo muy elevado para el ente rector.

¹⁴ Alternativamente, otros criterios de selección tradicionales de la estimación por Máxima Verosimilitud son implementables, ej. criterios de información AIC-BIC, criterios de ajuste como Pseudo R^2 , etc. Partiendo de que nuestro interés es la predicción del agregado laboral, favorecemos el criterio propuesto porque permite comparar directamente las predicciones con la tasa nacional de desempleo observada.

El recuento de individuos (*headcount*) para el cálculo de la tasa de desempleo se basa en la estimación de diversas especificaciones, en donde a cada individuo se lo reclasifica empleando múltiples cortes de probabilidad; por ejemplo, si el corte de probabilidad en evaluación es 0,7 y la probabilidad estimada del individuo de ser inactivo es $\hat{p}_i = 0,71$; el individuo pasa a contabilizarse directamente en esta categoría; es decir, sale de la PEA. De lo contrario (ej. $\hat{p}_i = 0,5$) se lo considera como parte de la PEA; y por tanto se lo clasifica como desempleado. Se calcula la tasa nacional de desempleo empleando los criterios y metodologías estándar aplicadas por el ente rector, ponderando las observaciones por los factores de expansión poblacional correspondientes. Finalmente, se compara la tasa nacional de desempleo pronosticada con la observada (calculada por el INEC). Se realiza este proceso iterativamente para diversos cortes y modelos (más detalles en la sección de Resultados). A menos que otro criterio de análisis particular esté explícito, la selección del modelo óptimo toma en cuenta el procedimiento descrito.

4.3. Ajustes adicionales: Reclasificación hacia adelante

Además del ejercicio de estimación propuesto mediante el modelo probabilístico, que incluye la corrección temporal mediante los efectos fijos en el tiempo (estimación estática)¹⁵, implementamos una estimación secuencial que reestima los parámetros del modelo seleccionado de manera iterativa. Esta estrategia incorpora, para la estimación en cada año, el ajuste del ejercicio de reclasificación del año previo según el modelo de referencia, en lugar de separar las muestras para las estimaciones dentro de la muestra (control) y las estimaciones fuera de la muestra (tratamiento). En otras palabras, una vez escogido el mejor modelo para el período 2000-2006, según el criterio de selección propuesto (menor MSE), empleando los parámetros obtenidos procedemos al pronóstico y reclasificación de los individuos del año 2007, en donde se calcula la tasa de desempleo correspondiente. Para el año 2008, reestimamos los parámetros del modelo incorporando los individuos reclasificados en el 2007; con el modelo resultante procedemos nuevamente con la reclasificación y posterior cálculo de la tasa de desempleo. Procedemos iterativamente en esta lógica hasta el año 2016.

Como se espera, este ejercicio de reclasificación es más conservador, porque incorpora estimaciones secuenciales, lo que conlleva, en el largo plazo, a una estimación más aproximada a la dinámica de la serie en la metodología nueva, en lugar de asumir una tendencia promedio de ajuste (modelo de efectos fijos temporales).

Finalmente, ambas aproximaciones ofrecen un intervalo para el pronóstico y, por tanto, se procede a obtener la estimación promedio como referencia para el análisis.

¹⁵ Esta denominación es (irremediablemente) imprecisa. Denominamos “estática” a esta aproximación no en el sentido de que no considera el factor temporal, sino en el sentido de que el pronóstico se realiza en función del efecto temporal promedio de la tendencia. La aproximación secuencial es “dinámica” en el sentido de que incorpora iterativamente las simulaciones previas para la reparametrización de los modelos de pronóstico, en cada año fuera de la muestra.

4.4. Datos

Los datos utilizados corresponden a la Encuesta Nacional de Empleo, Desempleo y Subempleo del Ecuador-ENEMDU, cuyo levantamiento está a cargo del INEC. Empleamos únicamente los levantamientos del último trimestre (mes de diciembre) para el periodo 2000-2016, con excepción del año 2002 por no disponer de información del sector rural en las bases empleadas para el estudio¹⁶. Debido a que el interés fundamental es la estimación de la población desempleada a partir de la reclasificación entre desempleados e inactivos, trabajamos exclusivamente con la información de los individuos sin empleo, con una muestra final de aproximadamente 235 mil observaciones para la reclasificación hacia atrás y 145 mil observaciones para la reclasificación hacia adelante.

5. RESULTADOS Y ANÁLISIS

5.1. Reclasificación hacia atrás

La Tabla 2 expone los resultados de las estimaciones del logaritmo de la razón de probabilidades (ecuación 6) de las diversas especificaciones analizadas. Como se anticipa, los coeficientes de las variables que tradicionalmente comprenden el análisis de la oferta laboral –y de capital humano– presentan signos negativos: años de escolaridad, edad y condición de jefe de hogar; es decir, tienen una relación inversa con la probabilidad de que el individuo sea reclasificado como inactivo. Por otro lado, los modelos del 4 al 6 corresponden a la misma especificación que los modelos del 1 al 3, excepto que en estos se incluyen las interacciones que permiten mejorar la precisión en la reclasificación al capturar la heterogeneidad de la interacción entre los años de escolaridad y las variables de género, área y ciudad. Por ejemplo, la interacción de años de escolaridad y la variable *dummy* para mujer presenta un signo negativo; en otras palabras, mientras más años de escolaridad tiene una mujer desempleada disminuye la probabilidad de ser reclasificada como inactiva. Alternativamente, se pueden analizar los efectos marginales de estas estimaciones buscando darle una explicación a los regresores; no obstante, a causa de que el interés de este documento está en el pronóstico de la tasa nacional de desempleo y no en la significancia y magnitud de los efectos de las variables incluidas, nos abstraemos de mayores conclusiones y argumentos en esta línea, recordándole al lector que la especificación propuesta no pretende una argumentación causal por la potencial violación de los supuestos discutidos en la sección metodológica.

La Tabla 3 muestra las mejores aproximaciones a la tasa nacional de desempleo, de acuerdo con los diferentes puntos de corte de la simulación considerados en cada modelo, para el periodo 2007-2016. La Figura 2 muestra la dinámica de la homologación para cada especificación, según el punto de corte óptimo seleccionado. El punto de corte es un valor definido entre cero y uno. Si la

¹⁶ Lo que conlleva una sobreestimación generalizada de los parámetros, por lo que se excluyó la información de este año y se procedió a emplear interpolaciones lineales respecto de los resultados del ejercicio propuesto.

TABLA 2
MODELO LOGIT PARA IDENTIFICACIÓN DE DESEMPLEADOS SUSCEPTIBLES DE
SER RECLASIFICADOS COMO INACTIVOS
 (reclasificación *hacia atrás*)

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
1 (inactivos)						
Escolaridad	-0,135*** (0,003)	-0,141*** (0,003)	-0,142*** (0,003)	-0,120*** (0,005)	-0,127*** (0,005)	-0,129*** (0,005)
Edad (Años)	-0,180*** (0,004)	-0,181*** (0,004)	-0,181*** (0,004)	-0,179*** (0,004)	-0,179*** (0,004)	-0,179*** (0,004)
Edad al cuadrado	0,003*** (0,000)	0,003*** (0,000)	0,003*** (0,000)	0,003*** (0,000)	0,003*** (0,000)	0,003*** (0,000)
Asistencia escolar=1	2,902*** (0,035)	2,888*** (0,035)	2,885*** (0,035)	2,881*** (0,035)	2,867*** (0,035)	2,866*** (0,035)
Mujer=1	1,165*** (0,023)	1,149*** (0,023)	1,155*** (0,023)	1,538*** (0,055)	1,526*** (0,056)	1,522*** (0,056)
Edad del jefe de hogar	0,046*** (0,003)	0,031*** (0,003)	0,019*** (0,003)	0,046*** (0,003)	0,031*** (0,003)	0,020*** (0,003)
Educación del jefe de hogar (años)	0,010*** (0,001)	0,008*** (0,001)	0,007*** (0,001)	0,010*** (0,001)	0,008*** (0,001)	0,007*** (0,001)
Jefe de hogar=1	-0,338*** (0,129)	-0,397*** (0,125)	-0,391*** (0,124)	-0,356*** (0,129)	-0,415*** (0,125)	-0,410*** (0,124)
Pareja=1	1,391*** (0,127)	1,413*** (0,122)	1,403*** (0,122)	1,376*** (0,127)	1,396*** (0,123)	1,386*** (0,122)
Hijo(a)=1	-0,205 * (0,119)	-0,168 (0,115)	-0,175 (0,115)	-0,203 * (0,119)	-0,166 (0,116)	-0,174 (0,115)
Yerno/Nuera=1	0,548*** (0,129)	0,548*** (0,125)	0,543*** (0,125)	0,543*** (0,129)	0,542*** (0,126)	0,538*** (0,125)
Nieto(a)=1	-0,320** (0,128)	-0,268** (0,126)	-0,271** (0,128)	-0,315** (0,128)	-0,264** (0,126)	-0,267** (0,125)
Padres o suegros=1	0,672*** (0,208)	0,712*** (0,206)	0,698*** (0,205)	0,640*** (0,209)	0,680*** (0,207)	0,664*** (0,206)
Otros parientes=1	-0,182 (0,126)	-0,148 (0,122)	-0,152 (0,122)	-0,179 (0,126)	-0,146 (0,123)	-0,150 (0,122)
Número de niños de 0 a 6 años	0,042*** (0,014)	0,105*** (0,014)	0,105*** (0,014)	0,042*** (0,014)	0,105*** (0,014)	0,105*** (0,014)
Número de ocupados en el hogar	0,157*** (0,011)	0,148*** (0,012)	0,140*** (0,012)	0,155*** (0,011)	0,146*** (0,012)	0,138*** (0,012)
Área rural=1	0,320*** (0,025)	0,410*** (0,025)	0,403*** (0,026)	0,172*** (0,062)	0,247*** (0,064)	0,182*** (0,065)
Subsidio=1	0,656*** (0,038)	0,678*** (0,038)	0,672*** (0,039)	0,636*** (0,039)	0,658*** (0,039)	0,656*** (0,039)
Quintil de ingresos q2=1		0,486*** (0,031)	0,492*** (0,031)		0,481*** (0,031)	0,486*** (0,031)
Quintil de ingresos q3=1		0,525*** (0,031)	0,531*** (0,031)		0,524*** (0,031)	0,528*** (0,031)
Quintil de ingresos q4=1		0,555*** (0,033)	0,546*** (0,033)		0,557*** (0,033)	0,547*** (0,033)
Quintil de ingresos q5=1		1,040*** (0,034)	1,018*** (0,034)		1,041*** (0,033)	1,019*** (0,034)
Nivel socioeconómico q2=1			-0,071 * (0,039)			-0,086** (0,040)
Nivel socioeconómico q3=1			0,068 (0,050)			0,060 (0,051)
Nivel socioeconómico q4=1			0,393*** (0,075)			0,375*** (0,074)
Nivel socioeconómico q5=1			0,113 (0,288)			0,104 (0,286)
Años de escolaridad*Mujer				-0,036*** (0,005)	-0,037*** (0,005)	-0,036*** (0,005)
Años de escolaridad*Área rural				0,016*** (0,006)	0,017*** (0,006)	0,023*** (0,006)

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Constante	2,600*** (0,205)	2,214*** (0,207)	2,422*** (0,209)	2,449*** (0,208)	2,064*** (0,210)	2,297*** (0,213)
Interacciones	No	No	No	Sí	Sí	Sí
Pseudo R-cuadrado	0,30	0,31	0,31	0,30	0,31	0,31
AIC	87457,99	86254,11	86179,47	87376,52	86168,99	86092,98
BIC	88017,86	86855,45	86822,28	87988,23	86822,17	86787,63
Clasificación correcta (%) (0,700)	92,33	92,49	92,47	92,27	92,44	92,43
Clasificación correcta (%) (0,703)	92,30	92,44	92,44	92,24	92,40	92,40
Clasificación correcta (%) (0,705)	92,27	92,42	92,43	92,21	92,37	92,37
Clasificación correcta (%) (0,708)	92,23	92,40	92,39	92,16	92,33	92,35
Clasificación correcta (%) (0,710)	92,20	92,37	92,36	92,13	92,30	92,31
Observaciones	235,131	235,131	235,131	235,131	235,131	235,131

Fuente: Encuesta Nacional de Empleo y Desempleo (INEC), todos los años.

Notas: Cada columna corresponde a una diferente regresión y cada una incluye una *dummy* de cohorte x año y cohorte x ciudad. Las interacciones son entre años de escolaridad y las variables *dummy* mujer, área rural y las ciudades de Guayaquil, Quito y Cuenca. La tabla reporta los coeficientes estimados del logaritmo de la razón de probabilidades de que un individuo desempleado sea reclasificado como inactivo. La clasificación correcta reporta el porcentaje de predicciones adecuadamente clasificadas en cada modelo cuando se clasifica como inactivos a aquellas observaciones cuya probabilidad estimada es mayor a 0,70, 0,703, 0,705, 0,708 y 0,71, respectivamente.

Errores estándar robustos agrupados a nivel del hogar en paréntesis.

* Nivel de significancia al 10%.

** Nivel de significancia al 5%.

*** Nivel de significancia al 1%.

probabilidad estimada de pertenecer al grupo de inactivos es mayor al valor del punto de corte, entonces se clasifica al individuo dentro de esta categoría (el valor de uno) y pasa a formar parte del conteo de individuos en esta condición¹⁷. Para implementar el criterio de selección óptima propuesto, aplicamos diferentes puntos de corte en el intervalo (0.5, 0.75). La Figura 3 presenta la dinámica en el período de análisis de la serie empalmada (homologada) versus la serie bruta, a partir del mejor punto de corte en el sentido de MSE, que se encuentra en 0,705. En otras palabras, a partir de la estimación realizada, la tasa nacional de desempleo se calcula mediante el conteo de los individuos cuya probabilidad de ser desempleado y ser reclasificado como inactivo es mayor a 0,705.

El modelo cuyo Error Cuadrático Medio de la Predicción es el menor, para el período de control (del 2007 a 2016), corresponde al modelo saturado (Modelo 6). No obstante, para el ejercicio de empalme este modelo presenta algunas dificultades; fundamentalmente, varios formularios de encuesta carecen de alguna información socioeconómica empleada en la especificación del modelo, para varios años previos al cambio metodológico. Como resultado, la predicción con este modelo es limitada, ya que ignora la influencia de las variables no disponibles para la estimación fuera de la muestra, lo que afecta directamente al error de predicción. En su lugar, siguiendo el criterio de selección propuesto, identificamos al Modelo 4 como óptimo para el pronóstico. Aunque relativamente

¹⁷ Por defecto, los *softwares* estadísticos suelen establecer un punto de corte específico para la reclasificación de la probabilidad, tradicionalmente si el pronóstico es mayor a 0,5; no obstante, este valor no responde a ningún criterio de optimalidad.

TABLA 3
VALORES ESTIMADOS POR ESPECIFICACIÓN Y PUNTOS DE CORTE
 (reclasificación *hacia atrás*)

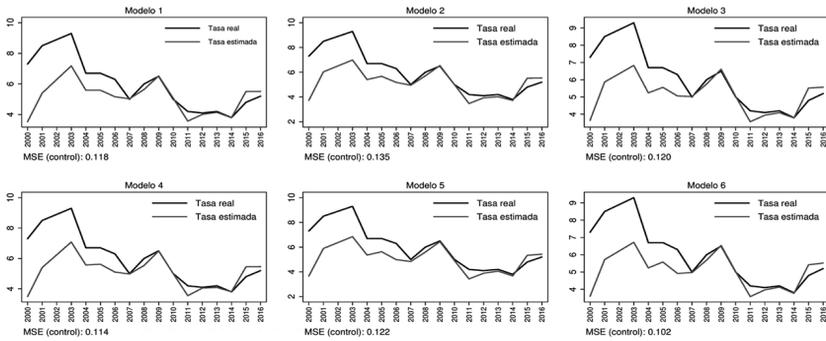
Años	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5		Modelo 6		Dif.* Real-Est. (%)				
	Puntos de corte																
	0,708	0,710	0,712	0,705	0,708	0,710	0,705	0,708	0,710	0,700	0,703	0,705		0,702	0,703	0,705	0,707
2000	3,5	3,6	3,6	3,7	3,8	3,6	3,6	3,7	3,5	3,5	3,7	3,7	3,7	3,6	3,6	3,6	3,7
2001	5,4	5,4	5,5	6,0	6,0	5,8	5,8	5,9	5,3	5,4	5,5	5,9	5,9	5,7	5,7	5,8	2,8
2003	9,3	7,2	7,3	6,9	7,0	6,7	6,8	6,8	7,0	7,1	7,1	6,8	6,9	6,7	6,7	6,8	2,6
2004	6,7	5,6	5,7	5,4	5,4	5,2	5,2	5,2	5,6	5,6	5,6	5,4	5,4	5,2	5,2	5,3	1,5
2005	6,7	5,6	5,7	5,6	5,7	5,5	5,5	5,6	5,6	5,6	5,7	5,6	5,7	5,5	5,6	5,6	1,1
2006	6,3	5,2	5,3	5,1	5,2	4,9	5,0	5,1	5,1	5,1	5,1	5,0	5,1	4,9	4,9	5,0	1,4
2007	5,0	5,0	5,1	4,9	5,0	4,9	5,0	5,0	4,9	5,0	5,0	4,8	4,9	4,9	5,0	5,0	0,0
2008	6,0	5,6	5,7	5,7	5,7	5,6	5,7	5,8	5,5	5,5	5,6	5,6	5,6	5,6	5,7	5,7	0,3
2009	6,5	6,6	6,6	6,5	6,6	6,5	6,6	6,6	6,4	6,5	6,5	6,4	6,4	6,5	6,5	6,6	0,0
2010	5,0	5,1	5,2	4,9	5,0	5,0	5,0	5,0	5,0	5,0	5,1	4,9	4,9	5,0	5,0	5,1	0,0
2011	4,2	3,6	3,7	3,4	3,5	3,5	3,5	3,6	3,5	3,6	3,6	3,4	3,4	3,5	3,6	3,6	0,6
2012	4,1	4,0	4,1	3,9	4,0	3,9	3,9	4,0	4,0	4,1	4,1	3,9	3,9	3,9	4,0	4,0	0,1
2013	4,2	4,1	4,2	4,0	4,1	4,0	4,1	4,1	4,0	4,1	4,2	4,0	4,1	4,1	4,1	4,2	0,0
2014	3,8	3,8	3,9	3,7	3,7	3,7	3,8	3,8	3,7	3,8	3,9	3,6	3,7	3,7	3,8	3,8	0,0
2015	4,8	5,5	5,6	5,5	5,6	5,4	5,5	5,5	5,4	5,5	5,5	5,3	5,4	5,4	5,4	5,5	-0,6
2016	5,2	5,5	5,6	5,5	5,5	5,5	5,5	5,6	5,4	5,5	5,5	5,4	5,4	5,5	5,5	5,5	-0,3
MSE (control)	0,118	0,118	0,127	0,137	0,135	0,127	0,124	0,120	0,127	0,117	0,114	0,134	0,122	0,109	0,102	0,108	

Fuente: ENEMDU (INEC), todos los años.

Notas: MSE-Mean Square Error (Error Cuadrático Medio de la Predicción). Cálculos incluyen todos los decimales, no mostrados en la tabla.

* Diferencia entre los valores observados y el pronóstico a base del modelo seleccionado (Modelo 4).

FIGURA 2
ESTIMACIÓN DE LA TASA NACIONAL DE DESEMPLEO (%), PARA CADA MODELO (reclasificación *hacia atrás*)

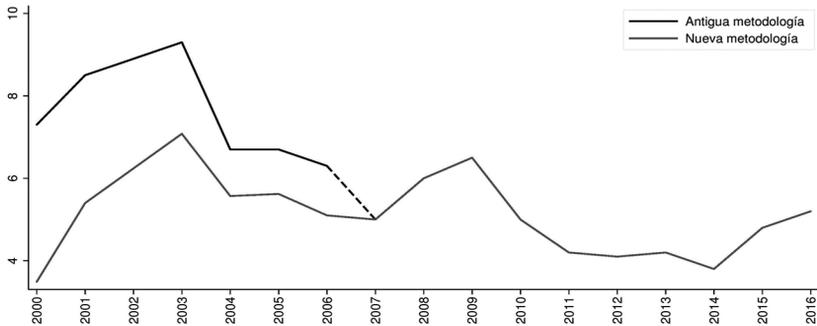


Fuente: ENEMDU (INEC), todos los años.

Nota: Para el año 2002, los valores fueron calculados por interpolación lineal.

Elaboración: Autores.

FIGURA 3
TASA NACIONAL DE DESEMPLEO (%), EMPALME (reclasificación *hacia atrás*)



Fuente: ENEMDU (INEC), todos los años; Sistema de Indicadores Sociales del Ecuador (SIISE).

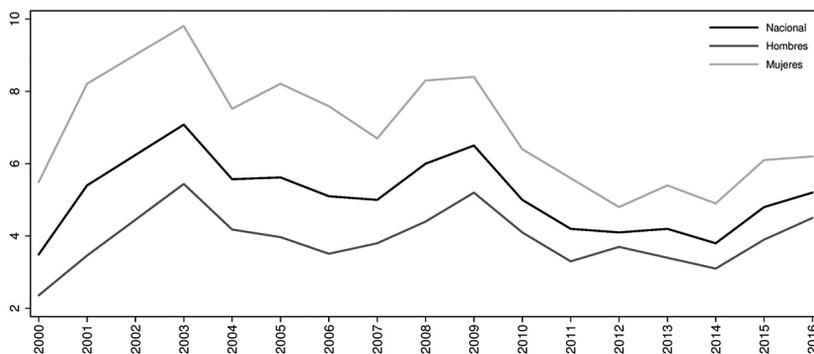
Nota: Para el año 2002, los valores fueron calculados por interpolación lineal.

Elaboración: Autores.

más conservador, en cuanto a su composición y especificación se priorizó tanto su nivel de precisión como la replicabilidad por la información disponible.

Como un ejercicio informativo adicional, una vez empalmada la tasa nacional de desempleo es posible desagregarla por género. Para ello se calcula la relación entre la tasa real de desempleo femenina/masculina respecto de la nacional, para el periodo 2000-2006, y luego se multiplican estos valores para cada tasa de desempleo estimada en el mismo periodo. La Figura 4 muestra la dinámica de la serie empalmada por género para el periodo de análisis. Como se puede observar, las tasas de desempleo son considerablemente mayores para las mujeres y la brecha permanece abierta respecto del género opuesto.

FIGURA 4
TASA NACIONAL DE DESEMPLEO (%), POR GÉNERO
(reclasificación hacia atrás)



Fuente: ENEMDU (INEC), todos los años.

Elaboración: Autores.

Nota: Para el año 2002, los valores fueron calculados por interpolación lineal; para ilustrar este proceso, se consideran las tasas de desempleo estimadas en el 2001 y 2003. A nivel nacional, estas tasas son 5,7% y 6,6%, respectivamente, luego se procede a calcular la siguiente fórmula: $[6,2\% = 0,5 (6,6\% - 5,7\%) + 5,7\%]$. De manera similar, se obtienen las tasas de desempleo por género, siendo 4,5% (es decir, $[0,5 (5,4\% - 3,5\%) + 3,5\%]$) para los hombres y 9,0% (es decir, $[0,5 (9,8\% - 8,2\%) + 8,2\%]$) para las mujeres.

Finalmente, la tasa nacional de desempleo presenta un diferencial de aproximadamente 2 puntos porcentuales entre la antigua y nueva metodología, para el periodo 2000-2006, en donde la serie empalmada conserva la misma tendencia de largo plazo que la serie original, pero ajustada por la reclasificación de los desempleados como inactivos.

5.2. Reclasificación hacia adelante

Podemos evaluar la dinámica de la tasa nacional de desempleo aplicando un análisis similar, esta vez partiendo de la antigua metodología y aplicándola a los años posteriores al cambio metodológico.

La Tabla 4 expone los resultados de las estimaciones de las múltiples especificaciones analizadas. La Tabla 5 y la Figura 5 resumen el procedimiento de selección del modelo de reclasificación óptimo para esta aproximación. Las conclusiones y argumentos de selección son similares a la sección previa. Vale aclarar que la diferencia en el tamaño de la muestra responde a las observaciones eliminadas automáticamente por falta de información en alguna de las variables del modelo de referencia. Para mantener consistencia con las especificaciones de la aproximación hacia adelante, priorizamos emplear modelos similares a costa de menor número de observaciones. Asimismo, consistente con lo expuesto respecto del grado de incertidumbre en este procedimiento, se puede observar que el nivel de ajuste de los modelos es relativamente inferior (Pseudo R^2 , AIC, BIC). El nivel de precisión en la reclasificación simulada es menor para los distintos puntos de corte empleados.

TABLA 4
MODELO LOGIT PARA IDENTIFICACIÓN DE DESEMPLEADOS SUSCEPTIBLES
DE SER RECLASIFICADOS COMO INACTIVOS
 (reclasificación *hacia adelante*)

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
l(inactivos)				
Años de escolaridad	-0,114*** (0,003)	-0,116*** (0,003)	-0,097*** (0,006)	-0,100*** (0,006)
Edad (años)	-0,164*** (0,005)	-0,166*** (0,005)	-0,165*** (0,005)	-0,166*** (0,005)
Edad al cuadrado	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)	0,002*** (0,000)
Asistencia escolar=1	2,468*** (0,039)	2,455*** (0,039)	2,445*** (0,040)	2,432*** (0,040)
Mujer=1	0,983*** (0,026)	0,973*** (0,026)	1,233*** (0,054)	1,230*** (0,055)
Educación del jefe de hogar (años)	0,047*** (0,003)	0,041*** (0,003)	0,048*** (0,003)	0,041*** (0,003)
Edad del jefe de hogar	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)	0,006*** (0,001)
Jefe de hogar=1	-0,333*** (0,107)	-0,344*** (0,108)	-0,333*** (0,107)	-0,345*** (0,108)
Pareja=1	1,006*** (0,103)	1,001*** (0,104)	1,006*** (0,104)	1,001*** (0,104)
Hijo(a)=1	-0,040 (0,099)	-0,027 (0,100)	-0,041 (0,099)	-0,029 (0,100)
Yerno/Nuera=1	0,770*** (0,115)	0,757*** (0,116)	0,766*** (0,115)	0,752*** (0,116)
Nieto(a)=1	0,222* (0,117)	0,236** (0,118)	0,219* (0,117)	0,233** (0,118)
Padres o suegros=1	0,586*** (0,156)	0,602*** (0,156)	0,574*** (0,156)	0,589*** (0,156)
Otros parientes=1	0,189* (0,109)	0,200* (0,109)	0,190* (0,109)	0,201* (0,110)
Número de niños de 0 a 6 años	0,005 (0,013)	0,031** (0,013)	0,006 (0,013)	0,031** (0,013)
Número de ocupados en el hogar	0,066*** (0,010)	0,050*** (0,011)	0,065*** (0,010)	0,048*** (0,011)
Área rural=1	0,254*** (0,029)	0,293*** (0,029)	0,180*** (0,061)	0,210*** (0,062)
Subsidio=1	0,149*** (0,037)	0,178*** (0,037)	0,128*** (0,037)	0,157*** (0,037)
Quintil de ingresos q2=1		0,176*** (0,036)		0,173*** (0,036)
Quintil de ingresos q3=1		0,235*** (0,037)		0,232*** (0,037)
Quintil de ingresos q4=1		0,277*** (0,038)		0,278*** (0,038)
Quintil de ingresos q5=1		0,443*** (0,038)		0,446*** (0,038)
Años de escolaridad*Mujer			-0,029*** (0,006)	-0,030*** (0,006)
Años de escolaridad*Área rural			0,009 (0,007)	0,011 (0,007)
Constante	2,743*** (0,155)	2,648*** (0,156)	2,624*** (0,159)	2,530*** (0,160)
Interacciones	No	No	Sí	Sí
Pseudo R-cuadrado	0,24	0,24	0,24	0,24
AIC	69289,99	69113,69	69256,03	69077,39
BIC	69724,90	69588,13	69740,36	69601,26
Clasificación correcta (%) (0,705)	89,95	90,04	89,87	89,94
Clasificación correcta (%) (0,708)	89,89	89,97	89,80	89,87
Clasificación correcta (%) (0,710)	89,85	89,92	89,75	89,83

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Clasificación correcta (%) (0,715)	89,75	89,80	89,87	89,73
Clasificación correcta (%) (0,718)	89,68	89,76	89,80	89,66
Observaciones	144,977	144,977	144,977	144,977

Fuente: Encuesta Nacional de Empleo y Desempleo (INEC), todos los años.

Notas: Cada columna corresponde a una diferente regresión y cada una incluye una *dummy* de cohorte x año y cohorte x ciudad. Las interacciones son entre años de escolaridad y las variables *dummy* mujer, área rural y las ciudades de Guayaquil, Quito y Cuenca. La tabla reporta los coeficientes estimados del logaritmo de la razón de probabilidades de que un individuo desempleado sea reclasificado como inactivo. La clasificación correcta reporta el porcentaje de predicciones adecuadamente clasificadas en cada modelo cuando se clasifica como inactivos a aquellas observaciones cuya probabilidad estimada es mayor a 0,705, 0,708, 0,71, 0,715 y 0,718, respectivamente.

Errores estándar robustos agrupados a nivel del hogar en paréntesis.

* Nivel de significancia al 10%.

** Nivel de significancia al 5%.

*** Nivel de significancia al 1%.

TABLA 5
VALORES ESTIMADOS POR ESPECIFICACIÓN Y PUNTOS DE CORTE
(reclasificación *hacia adelante*)

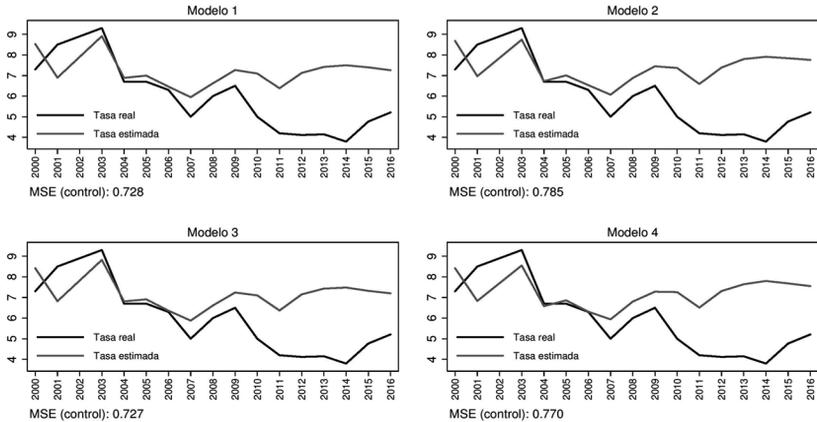
Años	Tasas observadas (%)	Modelo 1			Modelo 2			Modelo 3			Modelo 4			Dif.* Real-Est. (%)
		Puntos de corte												
		0,710	0,715	0,718	0,713	0,715	0,718	0,705	0,708	0,710	0,705	0,708	0,710	
2000	7,3	8,2	8,5	8,7	8,6	8,7	8,9	8,2	8,4	8,5	8,4	8,6	8,7	-1,1
2001	8,5	6,7	6,9	7,0	6,8	7,0	7,1	6,7	6,8	6,9	6,8	6,9	7,0	1,7
2003	9,3	8,6	8,9	9,0	8,7	8,7	8,9	8,7	8,8	8,9	8,6	8,7	8,8	0,5
2004	6,7	6,7	6,9	7,0	6,7	6,7	6,9	6,7	6,8	6,9	6,6	6,7	6,8	-0,1
2005	6,7	6,8	7,0	7,1	6,9	7,0	7,1	6,8	6,9	7,0	6,9	7,0	7,1	-0,2
2006	6,3	6,3	6,5	6,6	6,5	6,5	6,6	6,2	6,4	6,4	6,3	6,4	6,5	-0,1
2007	5,0	5,8	6,0	6,1	6,0	6,1	6,2	5,8	5,9	6,0	5,9	6,0	6,1	-0,9
2008	6,0	6,5	6,6	6,7	6,8	6,9	7,0	6,5	6,6	6,7	6,8	6,9	6,9	-0,6
2009	6,5	7,1	7,3	7,4	7,4	7,5	7,6	7,1	7,2	7,3	7,3	7,4	7,5	-0,7
2010	5,0	6,9	7,1	7,2	7,3	7,4	7,5	7,0	7,1	7,2	7,3	7,3	7,4	-2,1
2011	4,2	6,2	6,4	6,5	6,5	6,6	6,7	6,3	6,4	6,4	6,5	6,6	6,7	-2,2
2012	4,1	7,0	7,1	7,2	7,3	7,4	7,5	7,0	7,2	7,2	7,3	7,4	7,5	-3,0
2013	4,2	7,3	7,4	7,5	7,7	7,8	7,9	7,3	7,4	7,5	7,6	7,7	7,8	-3,3
2014	3,8	7,3	7,5	7,6	7,9	7,9	8,0	7,4	7,5	7,5	7,8	7,9	8,0	-3,7
2015	4,8	7,3	7,4	7,5	7,8	7,8	7,9	7,3	7,3	7,4	7,7	7,8	7,9	-2,6
2016	5,2	7,1	7,3	7,4	7,7	7,8	7,8	7,1	7,2	7,3	7,6	7,7	7,7	-2,0
MSE (control)		0,740	0,728	0,781	0,818	0,785	0,800	0,732	0,727	0,728	0,770	0,786	0,784	

Fuente: ENEMDU (INEC), todos los años.

Nota: MSE-Mean Square Error (Error Cuadrático Medio de la Predicción). Cálculos incluyen todos los decimales, no mostrados en la tabla.

* Diferencia entre los valores observados y el pronóstico a base del modelo seleccionado (Modelo 3).

FIGURA 5
ESTIMACIÓN DE LA TASA NACIONAL DE DESEMPLEO (%), PARA CADA MODELO
(reclasificación *hacia adelante*)



Fuente: ENEMDU (INEC), todos los años.

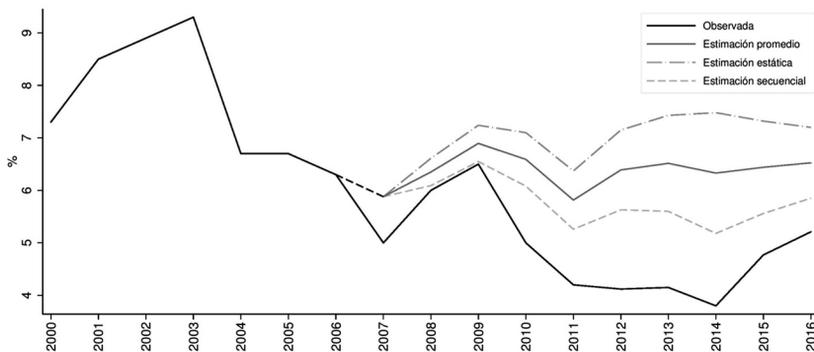
Nota: Para el año 2002, los valores fueron calculados por interpolación lineal.

Elaboración: Autores.

El modelo óptimo empleado para el pronóstico (estimación estática), equivalente a la metodología utilizada en la primera aproximación, es el Modelo 3, con un punto de corte de 0,708. Distinto a la reclasificación hacia atrás, y partiendo de los ajustes metodológicos correspondientes a base de la estimación secuencial (ver sección 4.3), la Figura 6 expone el resultado de los dos ejercicios de pronóstico realizados: estático y dinámico. Los resultados reafirman la dinámica observada en la primera aproximación de reclasificación (pronóstico hacia atrás) y, por tanto, exponemos las conclusiones del presente acápite sobre la base de la estimación promedio, como serie de referencia. A pesar de lo expuesto, vale recalcar que consideramos al intervalo entre la “estimación estática” y la “estimación dinámica” un espacio potencial de resultados y dejamos abierta la discusión para análisis y conclusiones alternativas.

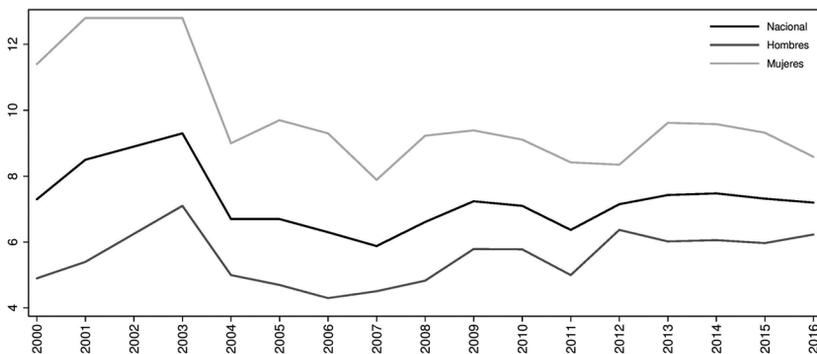
Como se puede observar en la Figura 6, la dinámica de la tasa de desempleo conserva la tendencia de la tasa observada (calculada por el ente rector); no obstante, distinto a lo que sucede en la estimación hacia atrás, en períodos específicos de la reclasificación hacia adelante existen divergencias relevantes. Mientras nuestro pronóstico predice una tendencia levemente al alza en la serie entre el 2012 y el 2013, la aplicación metodológica oficial muestra una dinámica relativamente constante. Asimismo, de manera significativa, mientras el ente rector reporta importantes incrementos en la tasa de desempleo a partir del 2014, nuestra aproximación predice una tendencia de crecimiento del desempleo más leve. Finalmente, por completitud exponemos el resultado del empalme de la serie por género en la Figura 7, empleando el mismo procedimiento descrito en la sección previa. La diferencia entre la estimación promedio de la tasa nacional de desempleo y la observada en la implementación de la nueva metodología por parte del ente rector es, en promedio, de aproximadamente 1,09 puntos porcentuales.

FIGURA 6
TASA NACIONAL DE DESEMPLEO
 (reclasificación *hacia adelante*)



Fuente: ENEMDU (INEC), todos los años.; Sistema de Indicadores Sociales del Ecuador (SIISE)
 Nota: Para el año 2002, los valores fueron calculados por interpolación lineal.
 Elaboración: Autores.

FIGURA 7
TASA NACIONAL DE DESEMPLEO (%) POR GÉNERO
 (reclasificación *hacia adelante*)



Fuente: ENEMDU (INEC), todos los años.
 Nota: Para el año 2002, los valores fueron calculados por interpolación lineal; para ilustrar este proceso, se consideran las tasas de desempleo 2001 y 2003. A nivel nacional, estas tasas son 8,5% y 9,2%, respectivamente, luego se procede a calcular la siguiente fórmula: $[8,9\% = 0,5 (9,2\% - 8,5\%) + 8,5\%]$. De manera similar, se obtienen las tasas de desempleo por género, siendo 5,7% (es decir, $[0,5 (6,0\% - 5,4\%) + 5,4\%]$) para los hombres y 13,4% (es decir, $[0,5 (14,0\% - 12,8\%) + 12,8\%]$) para las mujeres.
 Elaboración: Autores.

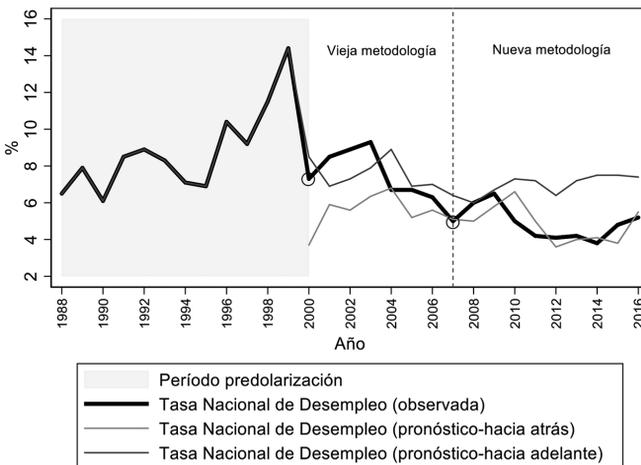
5.3. Análisis de estabilidad estructural de la serie de desempleo

El ejercicio propuesto está basado en el uso de información de carácter microeconómico, a nivel de individuos; por tanto, persiste la inquietud respecto de la independencia de la dinámica observada de la tasa de desempleo y la posi-

bilidad de cambios simultáneos de carácter estructural en la economía, más allá de los cambios metodológicos propuestos por la autoridad estadística. Por un lado, es posible que un individuo haya cambiado drásticamente su decisión de permanecer inactivo en el período de análisis. En otras palabras, independientemente del instrumento y metodología de medición, el umbral del salario de reserva del individuo pudo modificarse (ver modelo en sección 4.1.), y por tanto modificar la decisión de búsqueda de empleo, a la par con cambios en el salario de mercado. Analizar consistentemente la influencia de potenciales cambios en las preferencias de los agentes requiere de la identificación de shocks exógenos a nivel individuos/hogares, cuyas condiciones cumplan con los requerimientos fundamentales de un diseño de investigación (*research design*). La información disponible a nivel de las encuestas de empleo y desempleo, a nuestro mejor criterio, no permite el aprovechamiento de una estrategia de identificación adecuada para la construcción de un escenario contrafactual en nuestro análisis.

Por otro lado, a nivel agregado la economía pudo experimentar cambios en su estructura productiva durante el mismo período de modificación metodológica. Consecuentemente, *shocks* en las condiciones reales de la economía pueden intervenir en la reducción de los niveles de desempleo. Una forma de analizar si la reducción de las tasas de desempleo observadas se relaciona a cambios subyacentes en las relaciones económicas y no exclusivamente a los cambios en las mediciones de las estadísticas laborales, es evaluar la estabilidad o quiebre estructural, tanto de la serie de desempleo como de variables macroeconómicas relacionadas. Asimismo, este análisis permite evaluar si los pronósticos de las series ajustadas, propuestos en esta metodología, dan cuenta o corrigen la dinámica del quiebre estructural, condicional en su existencia.

FIGURA 8
ANÁLISIS DE QUIEBRE ESTRUCTURAL DE LA SERIE DE DESEMPLEO



Fuente: ENEMDU (SIISE-INEC), todos los años y estimaciones propias.

Nota: Para el año 2002, los valores fueron calculados por interpolación lineal.

Elaboración: Autores.

Una limitación del ejercicio de empalme de la serie de la tasa nacional de desempleo, propuesto en este trabajo, es la frecuencia de la información disponible. Al contar únicamente con una toma anual de la ENEMDU (diciembre de cada año), el número de observaciones para el análisis de series de tiempo es pequeño y el poder estadístico para las pruebas de estabilidad se reduce dramáticamente. Asimismo, la dinámica de todas las variables económicas en Ecuador sufre de un quiebre estructural marcado por la implementación del sistema de dolarización en enero del año 2000. Como muestra la Figura 8, este es también el caso de la tasa nacional de desempleo, cuya tendencia cambia dramáticamente a partir de este año (test F de quiebre estructural-Chow: p-value=0.0135). Esta realidad de las estadísticas económicas del país reduce el potencial análisis al uso de las observaciones del período posdolarización, para el que se genera el empalme de la serie de desempleo en este estudio.

Tomando en cuenta las limitaciones descritas procedemos, en primer lugar, a analizar la presencia de quiebre estructural en la serie de desempleo de Ecuador. El momento de interés para el quiebre (de fecha conocida) corresponde al año 2007, primer año de implementación de la nueva metodología. Como se mencionó previamente, la tasa de desempleo cayó en aproximadamente 1,3 puntos porcentuales este año y hay una diferencia de aproximadamente 2,6 puntos porcentuales en los niveles promedio del desempleo entre los períodos del 2000 al 2006, “vieja metodología”; y del 2007 al 2016, la “nueva metodología” (ver Figura 8).

Para implementar el test de estabilidad estructural de la serie de desempleo (tnd_t) se procedió a determinar el proceso estocástico subyacente. Mediante procedimientos estándar se determinó que se trata de un proceso autorregresivo de primer orden [AR(1)] y, por tanto, procedemos a evaluar la estabilidad del proceso empleando un modelo de estas características¹⁸. El modelo incorpora una variable binaria de quiebre estructural que representa la implementación de la “nueva metodología” y que toma un valor de 1 a partir del cambio metodológico, y cero para el período previo. La ecuación estimada es la siguiente:

$$(9) \quad tnd_t = \alpha_0 + \alpha_1 tnd_{t-1} + \delta(\text{nueva metodología}) + \epsilon_t$$

en donde ϵ_t corresponde a los errores ruido blanco del modelo. La Tabla 6 resume los resultados de este análisis para las tres series de la tasa nacional de desempleo: observada (calculada por el INEC), y las dos series empalmadas a base del pronóstico hacia atrás y hacia adelante, según la metodología descrita en este estudio.

El modelo propuesto hace un buen trabajo en describir la dinámica de la serie, por lo que procedemos directamente a la inferencia basados en los resultados expuestos en la Tabla 6¹⁹. La variable de quiebre estructural, “nueva

¹⁸ Un modelo de desempleo de estas características describe la “persistencia del desempleo” (Lindbeck, 1993). Alternativamente se probó un modelo de tendencia con *drift*. Todos los tests de especificación, ruido blanco y normalidad de los residuos, favorecen el proceso autorregresivo seleccionado.

¹⁹ Los residuos estimados de la regresión son estacionarios (carecen de raíz unitaria: ADF al 5% de significancia) y convergen a una distribución normal (test de normalidad sktest,

TABLA 6
ANÁLISIS DE QUIEBRE ESTRUCTURAL DE LA SERIE DE DESEMPLEO

	TND (observada)	TND (hacia adelante)	TND (hacia atrás)
Nueva metodología	-1,423* (0,782)	-0,867* (0,405)	-0,437 (0,344)
TND _{t-1} (correspondiente)	0,492* (0,259)	0,261 (0,204)	0,106 (0,235)
Constante	3,850* (1,890)	4,447*** (1,084)	6,665*** (1,698)
R-square	0,75	0,36	0,16
AIC	46,35	42,45	34,85
BIC	48,67	44,77	37,17
Observations	16,00	16,00	16,00

Fuente: ENEMDU (INEC), y estimaciones propias.

Notas: Errores estándar robustos en paréntesis.

* Nivel de significancia de 10%.

** Nivel de significancia de 5%.

*** Nivel de significancia de 1%.

metodología”, muestra evidencia débil de significancia estadística (nivel de significancia de 10%), tanto en la serie observada como en la serie empalmada hacia adelante. No obstante, la serie empalmada reduce el impacto del quiebre en aproximadamente 0,55 puntos porcentuales; en otras palabras, aun cuando existe evidencia débil de la presencia del quiebre en la serie empalmada, la magnitud del efecto se reduce en aproximadamente 40%. En el caso del empalme hacia atrás, el quiebre estructural desaparece.

Independientemente de las dudas respecto de la precisión de la inferencia, debido a los escasos grados de libertad disponibles, el empalme de la serie de desempleo contribuye a reducir el ruido generado por la implementación de los cambios metodológicos. En particular, contribuye a reducir la magnitud del quiebre estructural de la serie, lo que puede interpretarse como una recuperación de la dinámica de equilibrio.

En la misma línea de análisis, una alternativa complementaria es la de evaluar la independencia del potencial *shock* exógeno en otras variables macroeconómicas vinculadas a la dinámica del mercado del trabajo. Para ello, se puede analizar la presencia/ausencia de ruptura estructural al momento de implementación de los cambios metodológicos, esto es, el tercer trimestre de 2007. Desde el punto de vista macroeconómico, cambios en el nivel de desempleo de largo plazo están relacionados a varios factores, fundamentalmente a *shocks* de la oferta y demanda

Shapiro-Wilk y Shapiro-Francia: $pvalue > 10\%$, en todos los casos). Bajo estas condiciones, procedemos directamente a la inferencia estadística en función de las pruebas tradicionales (t y F) en función de errores robustos. Asimismo, no se evidencia presencia de correlación serial en los residuos; no obstante, la inferencia y conclusiones no se alteran si empleamos errores robustos a heterocedasticidad y autocorrelación (HAC). Resultados adicionales disponibles bajo solicitud.

del mercado. Teóricamente, la relación entre estas dos variables ha sido racionalizada mediante modelos de búsqueda (*search models*) (Lucas y Prescott 1974, Pissarides 1990, Aghion y Howitt 1992, 1994). La evidencia empírica, fundamentalmente en economías desarrolladas, favorece la existencia de una correlación negativa entre las tasas de crecimiento de la economía (medida por el PIB) y la tasa de desempleo; relación conocida como “Ley de Okun”²⁰. La Figura 9 resume la dinámica de las principales series macroeconómicas de las cuentas nacionales trimestrales de Ecuador: PIB, consumo de los hogares, consumo del Gobierno Central y la inversión (Formación Bruta de Capital Fijo-FBKF)²¹. Al tercer trimestre del 2007 no hay evidencia de ruptura en la dinámica de las series²², de hecho, no hay evidencia estadística de quiebre en ningún trimestre del año. La dinámica de las series macroeconómicas no presenta modificaciones o cambios representativos en períodos previos al cambio metodológico, lo que reduce la incertidumbre acerca de potenciales efectos dinámicos de *shocks* rezagados de las variables macroeconómicas en el nivel de desempleo. Independientemente de este análisis, la hipótesis de Okun sugiere una relación dinámica en donde la realización de la tasa de desempleo sea una función de rezagos de indicadores o variaciones de la actividad económica general. Este análisis está por fuera de los objetivos de este trabajo, ya que la frecuencia de la información para las series empalmadas impide su implementación; no obstante, condicional en la consistencia de las estadísticas laborales, objetivo central de este documento, constituye una interesante línea de investigación para futuros estudios.

5.4. Corolario: ¿cómo ha evolucionado el desempleo en Ecuador?

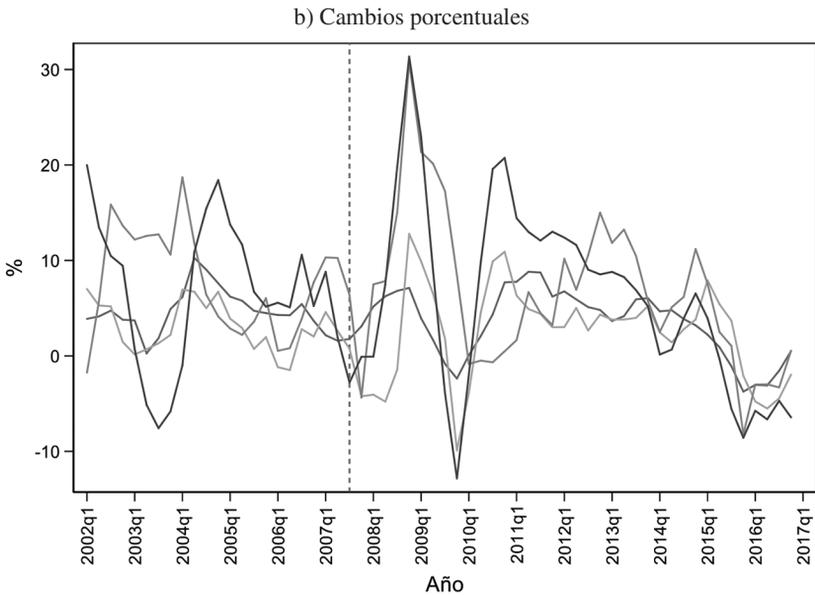
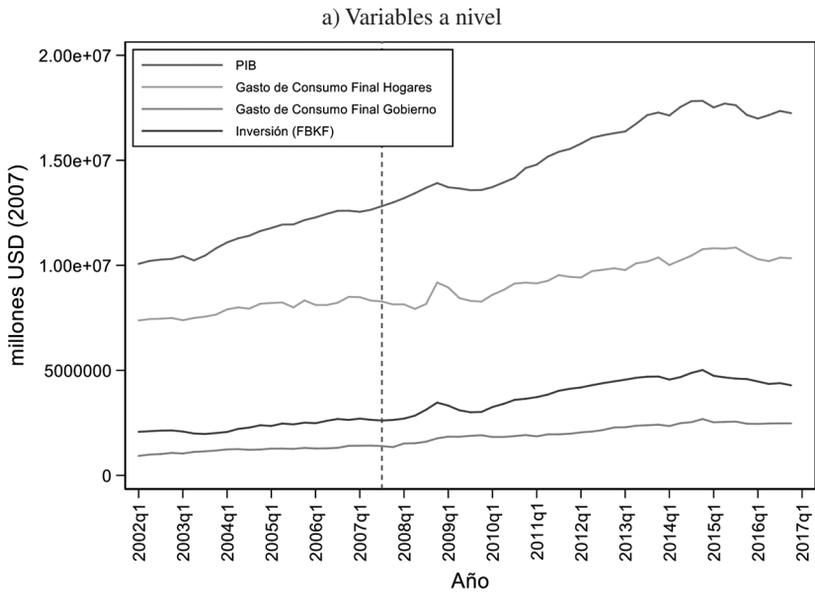
Una conclusión salta a la vista en ambas aproximaciones propuestas para el empalme de los indicadores laborales, en particular para la tasa nacional de desempleo en Ecuador: la dinámica del mercado laboral, en cuanto corresponde al nivel de desempleo, no se ha transformado de manera significativa y conserva los mismos niveles del período 2004 al 2006. Al 2016, el desempleo se encuentra en el mismo nivel que en el período de recuperación de la crisis financiera del país. Esto es un punto relevante de análisis y sugiere, al menos respecto de la dinámica en los niveles de desempleo, que las condiciones estructurales de generación y disponibilidad de empleo no se han dinamizado en casi 10 años a partir del cambio metodológico. Si bien se observa que la tasa de desempleo se redujo sistemáticamente a partir del año 2009, tanto en la serie observada

²⁰ No obstante, el argumento respecto de la veracidad de esta relación se mantiene en debate. Múltiples aplicaciones muestran una relación inestable, particularmente a nivel sub-nacional o regional (Moosa 1997, Villaverde y Maza 2009, Durech *et al.* 2014) y, por tanto, se considera más un argumento guía (*rule of thumb*) que una relación estructural. A pesar de sus debilidades, es un criterio que goza de credibilidad en los entornos de análisis macroeconómico y de predicción. Para una discusión más amplia ver Knotek (2007) y Ball *et al.* (2015).

²¹ Variables reales ajustadas en función del deflactor implícito del PIB, a precios del 2007. Información estadística del BCE.

²² Test de ruptura estructural estándar (Chow, Perron) realizados a partir del análisis de los procesos de cada serie, controlando por tendencia, estacionalidad y primeras diferencias (en el gráfico). Resultados disponibles bajo solicitud.

FIGURA 9
EVOLUCIÓN DE LAS SERIES MACROECONÓMICAS TRIMESTRALES
DE ECUADOR, PERÍODO 2000-2016



Fuente: Banco Central del Ecuador, Cuentas Nacionales Trimestrales.
 Elaboración: Autores.

como en las series empalmadas por ambas aproximaciones, al año 2011 esta dinámica se estabiliza. La crisis económica generada a raíz de la caída de los precios del barril de petróleo a fines del 2014, y que se evidencia en todos los indicadores macroeconómicos del país (ver Figura 9), transparentó la dinámica de la economía real en Ecuador. La dinámica de la capacidad de absorción del empleo por parte del sector privado se deterioró frente a una participación cada vez más activa del sector público, tanto en lo que se refiere a gasto corriente (ej. contratación de personal administrativo, sueldos y salarios de servidores públicos y fuerza pública) como a inversión.

¿Significa esto que no se han modificado las condiciones laborales generales en este período? Ciertamente no. Si bien los niveles de desempleo son similares entre los períodos pre y post-2007 (año del cambio metodológico), la tasa de desempleo constituye en sí mismo un indicador más resistente a cambios en la dinámica económica coyuntural. Por el contrario, es razonable pensar que gran parte de las transformaciones en el mercado laboral suceden en el subempleo, no en el desempleo y, por tanto, es posible que esta categoría haya sufrido transformaciones significativas en este período, transformaciones en las que no podemos establecer conclusiones en el presente estudio y deben formar parte de extensiones de esta investigación, que busquen analizar la dinámica del subempleo tanto en niveles como en composición (razones de subempleo), sectores y condiciones.

6. CONCLUSIONES

Los instrumentos estadísticos nacionales, como las encuestas laborales, están en permanente revisión, armonización y adaptación. Los entes rectores de las estadísticas nacionales están en la obligación de mejorar los mecanismos de recolección, análisis e interpretación de los datos; tanto para guiar la definición de estrategias de política pública como para informar el debate ciudadano respecto de la evolución de sus condiciones de bienestar. A pesar de ello, es consustancial a estos esfuerzos la generación de mecanismos de comparabilidad de la información, que permitan tener parámetros ciertos de referencia para analizar la evolución de las condiciones generales de la economía.

Este trabajo presenta un ejercicio de consolidación consistente para el empalme de las estadísticas laborales y constituye un esfuerzo por cubrir la carencia de información comparable en los indicadores del mercado laboral, fundamentales para el análisis económico y la evaluación de resultados de las políticas laborales. A partir de las experiencias internacionales, el presente trabajo presenta una propuesta de adaptación de la metodología de empalme de la serie de desempleo para el caso ecuatoriano. Innova al establecer criterios de simulación y selección específicos que permitan una estimación *óptima* a partir de la capacidad predictiva de modelos tradicionales de análisis. Asimismo, la aproximación expuesta mejora la capacidad de pronóstico, tanto desde el punto de vista de la información disponible como al considerar las dificultades y sesgos derivados del uso de información desagregada a nivel de encuestas de hogares.

Nuestros resultados favorecen el uso de información microeconómica de las encuestas de hogares para el empalme y armonización de las series temporales del mercado laboral. Esta aproximación es superior a otras alternativas de ca-

rácter agregado en donde el trabajo directo con series de tiempo desconoce las decisiones de oferta laboral que se desarrollan a nivel del hogar. Asimismo, la descomposición de los ciclos de las series temporales no ofrece una respuesta frente a los potenciales sesgos de agregación de la información, derivados fundamentalmente de cambios en las expectativas de los agentes. Los individuos enfrentan una dinámica económica a nivel de los determinantes laborales del hogar, que cambia en el tiempo, y la capacidad de captar información mediante los instrumentos estadísticos requiere ajustarse a este proceso.

En el caso particular de Ecuador, los resultados muestran que las condiciones estructurales de generación de empleo, evaluados por intermedio de la tasa nacional de desempleo en la economía, no se han modificado sustancialmente a partir del año de cambio metodológico, contradiciendo la tendencia observada por la evolución del indicador en las estadísticas oficiales, no comparable entre los períodos de análisis.

REFERENCIAS

- Aghion, P. and Howitt, P. (1992). "A Model of Growth Through Creative Destruction", *Econometrica*, 60, 323-351.
- Aghion, P. and Howitt, P. (1994). "Growth and Unemployment", *The Review of Economic Studies*, 61(3), 477-494.
- Arango-Thomas, L.; García-Suaza, A. y Posada, C. (2006). La metodología de la Encuesta Continua de Hogares y el empalme de las series del mercado laboral urbano de Colombia, *Banco de la República-Documentos de Trabajo*.
- Ball, L.; De Roux, N. and Hofstetter, M. (2013). "Unemployment in Latin America and the Caribbean", *Open Economies Review*, 24(3), 397-424.
- Ball, L.; Jalles, T. and Loungani, P. (2015). "Do Forecasters Believe in Okun's Law? An Assessment of Unemployment and Output Forecasts", *International Journal of Forecasting*, 31(1), 176-184.
- Banco Central del Ecuador. (2008). Indicadores del Mercado Laboral: La transición FLACSO-INEC. [PPT].
- Borjas, George (2016). *Labor Economics* (7th ed.), Boston: McGraw Hill.
- Correa, J. y Castro, J. (2006). Una propuesta de empalme de las encuestas de hogares basada en la tasa de desempleo, *Estudios Gerenciales*, Colombia.
- Durech, R.; Minea, A.; Mustea, L. and Slusna, L. (2014). "Regional evidence on Okun's Law in Czech Republic and Slovakia", *Economic Modeling*, 42, 57-65.
- Guataquí, J. and Tabora, R. (2006). "Theoretical and empirical implications of the new definition of unemployment in Colombia", *Revista de Economía del Rosario*, 9(1), 21-38.
- Instituto Nacional de Estadísticas y Censos (2011). Encuesta de Estratificación de Nivel Socioeconómico-NSE. Recuperado de: www.ecuadorencifras.gob.ec/documentos/web/inec/Estadisticas_Sociales/Encuesta_Estratificacion_Nivel_Socioeconomico/Cuestionario_Estratificacion.pdf.
- Instituto Nacional de Estadísticas y Censos (2014). Metodología para la medición del empleo en el Ecuador. *Informes Metodológicos*.

- Knotek, Edward S. (2007). "How Useful Is Okun's Law". *Economic Review-Federal Reserve Bank of Kansas City*, 92(4), 73-103.
- Lasso, Francisco (2002). *Nueva metodología de encuesta de hogares ¿Más o menos desempleados?*, Archivos de Economía N.213, Departamento Nacional de Planeación-Dirección de Estudios Económicos, Colombia.
- Lee, J. (2000). "The robustness of Okun's law: Evidence from OECD countries". *Journal of macroeconomics*, 22(2), 331-356.
- Lindbeck, Assar (1993). *Unemployment and Macroeconomics* (Vol. 3), The MIT Press.
- Lucas, R. and Prescott, E. (1974). "Equilibrium Search and Unemployment". *Journal of Economic Theory*, 7, 188-209.
- Moosa, I. (1997). "A Cross-Country Comparison of Okun's Coefficient", *Journal of Comparative Economics*, 24, 335-356.
- Pissarides, C. (1990) *Equilibrium Unemployment Theory*. Oxford: Basil Blackwell.
- Villaverde, J. and Maza, A. "The robustness of Okun's law in Spain, 1980-2004 Regional evidence", *Journal of Policy Modeling*, 31, 289-297.

