

## NOTA METODOLÓGICA

# **La metodología de la Encuesta Continua de Hogares y el empalme de las series del mercado laboral urbano de Colombia**

## *The methodology of the Continuous Household Survey and the connection of the Colombian series urban labor market*

Luis Eduardo Arango  
Andrés Felipe García  
Carlos Esteban Posada\*

### **Resumen**

La encuesta de hogares tuvo cambios importantes en el año 2000. Ello implicó modificaciones en los conceptos, preguntas, periodicidad de recolección y cobertura que han dificultado la realización de estudios sobre el mercado laboral que requieran series completas desde 1984 hasta hoy. Efectos del cambio son menores registros en las tasas de desempleo y participación y uno mayor en la tasa de ocupación. Este

---

\* Las opiniones contenidas en este documento son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva. Luis Eduardo Arango y Carlos Esteban Posada son integrantes de la Unidad de Investigaciones Económicas de la Gerencia Técnica y Andrés Felipe García, que se desempeñó como asistente de esta investigación, es estudiante de economía de la Universidad de Antioquia. Se agradecen los comentarios de Francisco Lasso y las sugerencias y recomendaciones de dos evaluadores anónimos. Las series estimadas están disponibles con la solicitud previa a los autores.

Este artículo fue recibido 8 de junio de 2007, modificado el 22 de abril de 2008 y aceptado el 30 de mayo siguiente.

artículo expone dos métodos de empalme de las series: uno apoyado en datos micro y otro en agregados del mercado laboral. Con base en los resultados de ambos métodos, proponemos series de tasas de ocupación, participación y desempleo para empalmar con las series recientes del Dane generadas a partir de la Encuesta Continua de Hogares.

*Palabras clave:* fuerza laboral, encuesta de hogares, desocupados, trabajadores familiares sin remuneración, modelos logit, modelos panel.

*Clasificación JEL:* J21, J22, J82, C21, C22, C23.

## **Abstract**

The Continuous Household Survey of Colombia had important changes in 2000 with respect to concepts, questions, frequency of data collection and sample covering. This has made difficult to carry out studies on the labor market that demand series from 1984 up until now. The effects of the changes mirror in the reduction of the unemployment and participation rates while the increase in the occupation rate. This article presents two forms of joining the series: one built on micro data and another based on the labor market. The results allow us to propose series of employment, unemployment and participation rates to join with the recent official series generated from the Continuous Household Survey.

*JEL classification:* J21, J22, J82, C21, C22, C23.

*Key words:* labor force, household survey, unemployment, family workers without pay, logit models, panel data model.

## **Introducción**

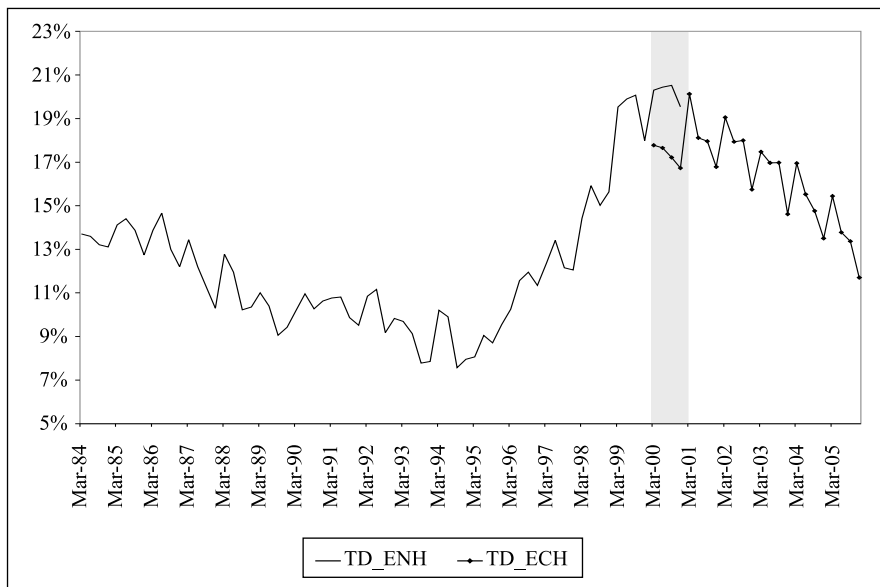
En el año 2000, el Dane realizó un profundo cambio en la metodología del sistema de encuestas de hogares con el fin de acoger los estándares internacionales surgidos de la Conferencia Internacional de Estadís-

ticos del Trabajo (CIET) de 1993<sup>1</sup>. Ello significó la interrupción de las series del mercado laboral colombiano derivadas de la Encuesta Nacional de Hogares (ENH).

El cambio se consolidó a partir del segundo trimestre de 2001, cuando se dio paso a un nuevo sistema de encuestas conocido como Encuesta Continua de Hogares (ECH), el cual tiene mayor cobertura, frecuencia en la recolección y presentación de datos (mensual en lugar de trimestral), además de cambios importantes en los conceptos que permiten estimar la composición de la fuerza de trabajo.

Con la ECH no es posible realizar análisis comparativos con el período previo a 2001, pues la caracterización de la fuerza de trabajo y la definición de los diferentes grupos de población tuvieron variaciones importantes, en especial en la definición de los desocupados, que afectaron los niveles de los principales agregados del mercado laboral, como se ilustra en el gráfico 1, en la cual se presentan las tasas desempleo (TD) en el período 1984-2005 calculadas con base en la ENH y la ECH.

**Gráfico 1. Tasa de desempleo total siete ciudades. 1984-2005.**



Fuente: ECH y ENH del Dane.

<sup>1</sup> LA CIET es una conferencia de expertos de la Organización Internacional del Trabajo (OIT) que se reúne desde 1923. Las conferencias claves para efectos metodológicos fueron

Se puede observar allí cómo en el año 2000 (único año en el que se tienen datos tanto para la ENH como para la ECH), según la nueva metodología, la tasa de desempleo es inferior en 2,8 puntos porcentuales (*pp*), en promedio, a la que se generaba según los criterios de la ENH<sup>2</sup>. Es claro entonces que para analizar la dinámica del mercado laboral en el mediano plazo es necesario empalmar las diferentes variables con el fin de mantener la unidad desde los puntos de vista metodológico y cuantitativo.

El objetivo de este documento es presentar propuestas de empalme de las series de los indicadores del mercado laboral, correspondientes a siete ciudades, para el período 1984-2005. Para ello, se utilizan dos enfoques. El primero –microeconómico–, basado en la ECH entre 2001 y 2005, utiliza modelos de elección binaria para pronosticar la probabilidad de ser reclasificado como inactivo o trabajador familiar sin remuneración (TFSR), según el caso. Este método se desarrolla a partir de los aportes de Lasso (2002). Sin embargo, dadas las dificultades con el ajuste de estos modelos para reclasificar los diferentes grupos de población, se emplea un segundo enfoque que se apoya en los agregados del mercado laboral. Este método se funda en el análisis de la tasa global de participación y los ajustes en las definiciones de la fuerza de trabajo principalmente en el grupo de los desocupados y el de trabajadores familiares sin remuneración.

Este documento tiene seis secciones, de las cuales esta introducción es la primera. En la segunda sección se hace una breve reseña de la transición de la ENH a la ECH con especial acento en el cambio en las definiciones de la fuerza laboral; en la tercera sección se discuten brevemente los empalmes pioneros de Suárez y Buriticá (2002) y Lasso (2002); la cuarta sección presenta la guía teórica y la estimación del

---

la de 1983, en la cual se establecieron algunos referentes de medición principalmente del subempleo; la de 1998, que se centró en el sector informal, y en 1998 fue el subempleo el concepto clave.

<sup>2</sup> Se debe tener en cuenta que en el año 2000 todavía no se habían realizado todos los cambios metodológicos; estos terminaron en junio de 2001. En consecuencia, la diferencia en la tasa de desempleo podría ser menor. Al comparar la tasa global de participación con la tasa de ocupación para el año 2000 según ENH y ECH, se tiene que la primera es menor, en promedio, en 0,58 *pp*, mientras que la tasa de ocupación es mayor, en promedio, en 1,42 *pp*.

modelo microeconómico; la quinta sección contiene el modelo basado en los agregados del mercado laboral y los resultados; finalmente, la sexta sección presenta algunas conclusiones.

## **I. Transición metodológica: de la ENH a la ECH**

Las encuestas de hogares, instrumento básico para estudiar la fuerza de trabajo en Colombia, se han venido utilizando desde los años sesenta. Entre 1970 y 1975, el Dane realizó nueve etapas que, aunque no tenían un esquema unificado, sirvieron de base para el desarrollo del ‘sistema de encuestas de hogares’, el cual se consolidó en 1976 bajo el nombre de Encuesta Nacional de Hogares (ENH). Su marco conceptual y sus definiciones estuvieron inspirados en resoluciones de la CIET y la OIT.

Hasta 1984, con base en la ENH, el Dane reportaba datos agregados trimestralmente de las cuatro principales ciudades (Bogotá, Medellín, Cali y Barranquilla) y anualmente para el total nacional y cabeceras de otras zonas. A partir de ese año la encuesta se amplió a siete ciudades: se sumaron Bucaramanga, Pasto y Manizales y sus áreas metropolitanas.

Sin embargo, la ENH presentaba algunas limitaciones, entre las que destacaban la frecuencia trimestral<sup>3</sup> y la gran rotación del personal encargado de capturar información, lo cual dificultaba los operativos de campo y el seguimiento a los hogares no informantes (Lasso, 2002). Con el objetivo, entonces, de modernizar el sistema de encuesta de hogares y obtener así resultados que reflejaran de mejor manera el mercado laboral y su dinámica, el Dane comenzó a desarrollar en 1996 un plan para mejorar la calidad de la ENH actualizando los marcos metodológico y de muestra, la optimización de los procesos operativos y la incorporación de tecnología para facilitar la planeación de los operativos de campo. La adopción de estos elementos trajo como resultado el “sistema continuo de recolección de datos”, el cual toma en cuenta los cambios metodológicos propuestos por la OIT en 1993, las

---

<sup>3</sup> Lo cual dejaba amplios períodos sin información que ignoraban aspectos relativos a la estacionalidad y movilidad laborales.

experiencias internacionales<sup>4</sup> y la dinámica del mercado laboral colombiano. Este sistema fue adoptado, de manera paulatina, a partir de enero de 2000 bajo el nombre de Encuesta Continua de Hogares (ECH), el cual además del contenido de la encuesta y su periodicidad de aplicación, requiere una logística y un operativo técnico muy importantes.

El uso de este nuevo enfoque implicó dos cambios importantes en las estadísticas del mercado laboral: las definiciones de fuerza de trabajo (ocupados, desocupados e inactivos) y, el sistema continuo de recolección<sup>5</sup>. El primer cambio es el de mayor magnitud, pues produce una modificación en la composición de la población en edad de trabajar (PET), lo cual provoca reducciones en los niveles de las tasas de desempleo y participación y un aumento en el de la tasa de ocupación. El segundo no afecta la medición de los indicadores sino la estructura serial de éstas, por los ajustes estacionales que se producen en los datos intratrimestrales. A continuación se hace un recuento de los principales cambios de orden conceptual recogidos en la ECH.

#### **A. El caso de los trabajadores familiares sin remuneración (TFSR)**

La ECH modificó la definición de los TFSR, uno de los componentes de la variable ‘ocupados’. Así, mientras que en la ENH se consideraban como tales las personas que trabajaban en un negocio familiar sin remuneración durante 15 ó más horas a la semana, en la ECH basta con trabajar al menos una hora semanal para ser considerado TFSR. En consecuencia, el grupo de TFSR que trabajaba más de una y menos de 15 horas semanales, que estaba clasificado como desocupado o inactivo (según el flujo de respuestas del módulo de fuerza de trabajo), es ahora reclasificado como ocupado.

El diagrama 1 muestra el flujo de preguntas que permiten identificar los TFSR. En primer lugar, las personas se clasifican según la actividad

---

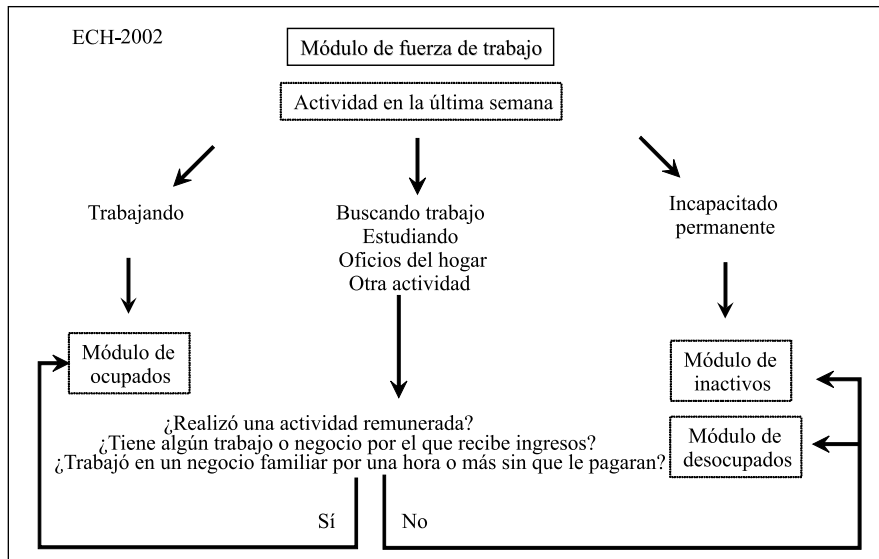
<sup>4</sup> A propósito de modificaciones Guataquí y Taborda (2006) muestran en qué consistieron los cambios metodológicos llevados a cabo en el Reino Unido, Trinidad y Tobago, España e Italia, cuyos resultados fueron las caídas en las tasas estimadas de desempleo en, aproximadamente, 2 *pp*, en promedio.

<sup>5</sup> Descrito por el Dane como una serie continuada de ‘ciclos’ de encuesta.

realizada durante la última semana (o semana de referencia para la encuesta); esto es, si estuvieron ocupadas (trabajando), desocupadas o inactivas (buscando trabajo, dedicadas a estudiar, a oficios del hogar o alguna otra actividad) o se declararon inactivas a causa de una incapacidad permanente. Dentro del grupo de las personas ocupadas, se clasifican como TFSR aquellas que trabajan en un negocio familiar sin recibir remuneración al menos una hora a la semana. El gráfico 2 muestra las personas clasificadas como TFSR que trabajan entre una y catorce horas a la semana como proporción del total de desocupados e inactivos. Se observa allí cómo la población reclasificada por este criterio, en las siete principales ciudades, equivale, en promedio, a 0,8% del total de desocupados e inactivos.

Este cambio metodológico trajo como resultado el aumento de la población ocupada y la disminución de los desocupados e inactivos; por tanto, disminuyeron los registros de las tasas de desempleo (TD)<sup>6</sup> y de participación (TGP) y aumentó el de la tasa de ocupación (TO).

Diagrama 1. Flujo de preguntas del módulo de fuerza de trabajo para la clasificación de TFSR.



Fuente: formularios ECH-Dane 2002.

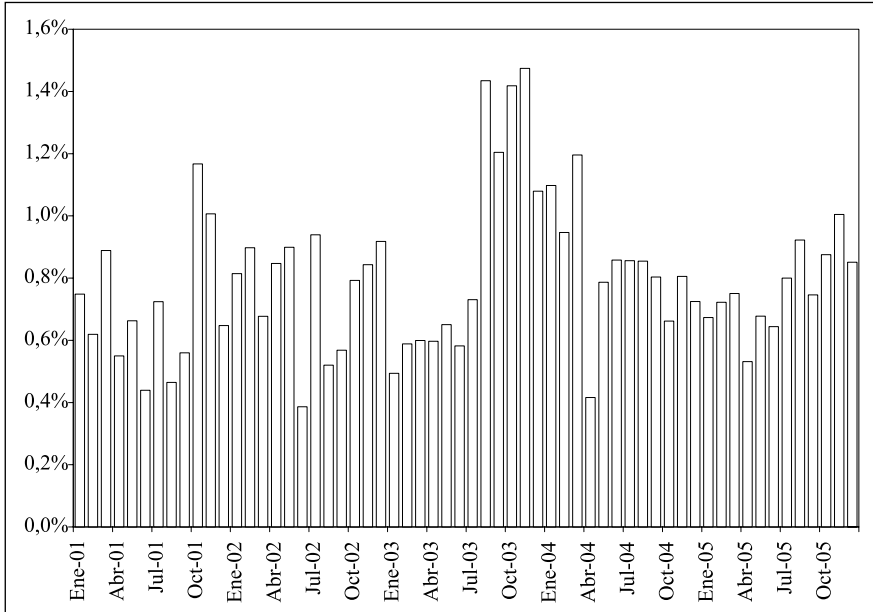
<sup>6</sup> Lo cual se explica gracias a que la población económicamente activa (PEA) aumenta, al menos en el caso de que un inactivo se clasifique como TFSR.

## B. El caso de los desocupados

La población de desocupados incluye dos grandes grupos: quienes están en situación de desempleo abierto y en situación de desempleo oculto. Para efectos de identificación y clasificación de unos y otros, la ECH introduce el concepto de *disponibilidad*. En este sentido, la ECH incluye las siguientes tres preguntas: *i) ¿... desea conseguir un trabajo remunerado o instalar un negocio?; ii) Aunque desea trabajar, ¿por qué motivo principal no hizo diligencias en las últimas cuatro semanas? (aparecen listadas las razones no válidas de desempleo); iii) Si le hubiera resultado algún trabajo... ¿estaba disponible la semana pasada para empezar a trabajar?*

Estas preguntas permiten capturar la información necesaria para clasificar la población desocupada bajo la nueva definición de desocupado de la ECH, la cual se resume y compara con la definición de la anterior metodología en el cuadro 1. Se observan allí dos cambios importantes; el primero, da cuenta del periodo de referencia, el cual cambia de una semana a cuatro; y el segundo incluye el factor de disponibilidad inmediata a trabajar y razones válidas de desempleo.

Gráfico 2. TFSR reclasificados como proporción del total de desocupados más inactivos. Siete ciudades. Enero 2001 - diciembre 2005.



Fuente: ECH-Dane. Cálculos de los autores.



Cuadro 1. Definiciones de lo que es un desocupado en la ENH y la ECH.

ENH	ECH
<b>Desempleo abierto:</b>	<b>Desempleo abierto:</b>
Población que estuvo sin empleo en la semana de referencia e hizo alguna acción para conseguir empleo en la semana de referencia.	Población que estuvo sin empleo en la semana de referencia. Realizó alguna acción para conseguir empleo en las últimas cuatro semanas y tenía disponibilidad inmediata para empezar a trabajar en la semana de referencia
<b>Desempleo oculto:</b>	<b>Desempleo oculto:</b>
Población que estuvo sin empleo en la semana de referencia y no hizo acción alguna para conseguir empleo en la semana de referencia pero sí en el último año.	Sin empleo en la semana de referencia,  No hizo acción alguna para conseguir empleo en las últimas cuatro semanas, pero sí en el último año.  Tenía razón válida de desempleo (en las últimas cuatro semanas) y,  Tenía disponibilidad inmediata para empezar a trabajar en la semana de referencia.

Fuente: Dane (2001) y Lasso (2002), pág. 17.

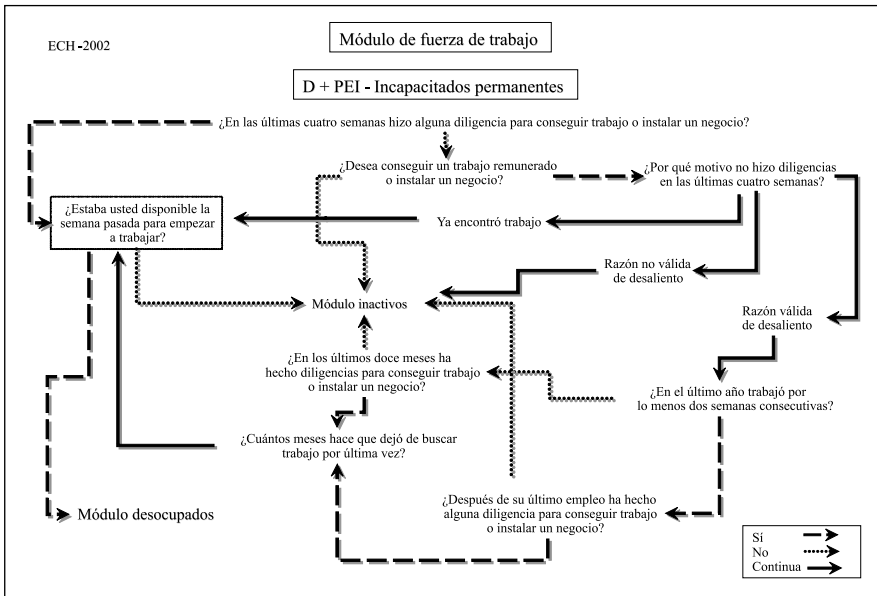
Por tanto, en comparación con la ENH no se consideran desocupadas en la ECH aquellas personas que se encuentran en situación de desempleo abierto pero no tienen disponibilidad inmediata para trabajar ni aquellas que están en situación de desempleo oculto pero que no cumplen con al menos una de las tres condiciones siguientes: *i*) tener disponibilidad inmediata para trabajar, *ii*) tener deseo de conseguir un trabajo remunerado o instalar un negocio, o *iii*) tener una razón válida de desempleo<sup>7</sup>.

A partir de la nueva definición de ‘desocupados’ se puede establecer un flujo de preguntas que permite determinar la magnitud del cambio en la definición de los desocupados que afecta, además de este grupo, a los inactivos, de tal forma que se observa una disminución de la medida de los desocupados en contraste con un aumento de la medida

<sup>7</sup> Se consideran razones no válidas de desempleo las siguientes: considerarse muy joven o muy viejo, tener responsabilidades familiares, problemas de salud, estar estudiando o tener otra razón diferente. Las razones válidas de desempleo son: haber encontrado trabajo, no hay trabajo disponible en la ciudad, estar esperando a que lo llamen, no saber cómo buscarlo, estar cansado de buscar, no encontrar trabajo en su oficio o profesión, estar esperando la temporada alta, carecer de la experiencia necesaria, no tener recursos para instalar un negocio, o los empleadores lo consideran muy viejo o muy joven.

de la población de inactivos. Dicho flujo se presenta en el diagrama 2, el cual se construyó con base en la ECH de 2002, pues entre 2000 y el primer semestre de 2001 las preguntas del módulo de fuerza de trabajo aún eran objeto de modificaciones. Sin embargo, ello no impidió, para estos períodos, caracterizar los desocupados bajo la definición presentada en el cuadro 1.

Diagrama 2. Flujo de preguntas del módulo de fuerza de trabajo para la clasificación de desocupados e inactivos.



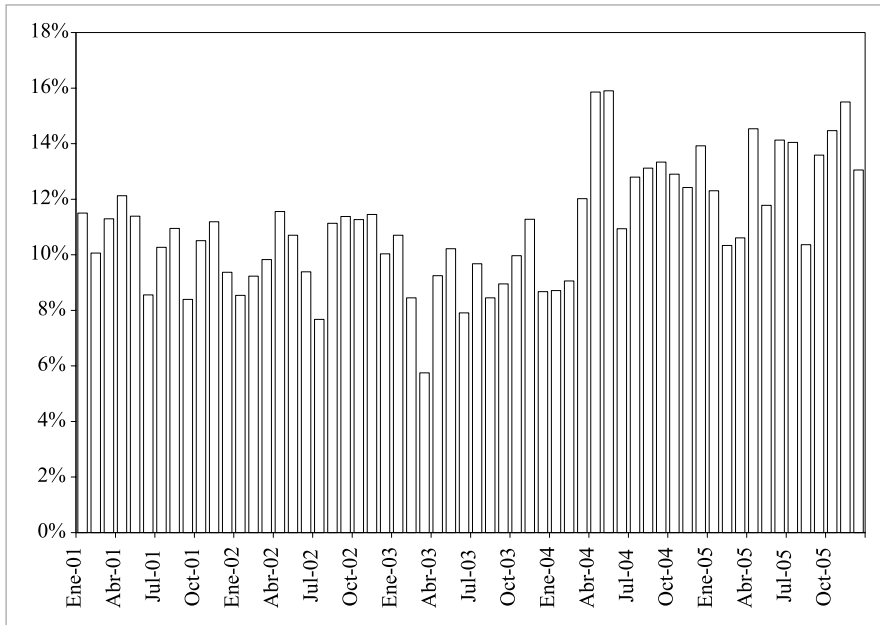
Fuente: Formulario ECH 2002.

El diagrama 2 parte de los grupos aún sin clasificar, es decir, la población en edad de trabajar, sin contar los ocupados (incluso los TFSR) y los inactivos por razones de incapacidad permanente. De este grupo se clasifican en primer lugar los desempleados en condición de desempleo abierto según las características que presenta el cuadro 1 (sin trabajo en la semana de referencia, buscó trabajo en las últimas cuatro semanas y estaba disponible para trabajar inmediatamente), seguidos de los desempleados en condición de desempleo oculto (según los criterios de razón válida de desempleo y disponibilidad a trabajar); el grupo restante se clasifica como inactivo.

Para ilustrar la magnitud del efecto que significó este cambio metodológico, el gráfico 3 muestra el porcentaje de desocupados que

son reclasificados como inactivos en la ECH entre enero de 2001 y diciembre de 2005 correspondientes a siete ciudades. Entre enero de 2001 y diciembre de 2003 el promedio de reclasificados es de 9,9%, y entre enero de 2004 y diciembre de 2005, de 12,7%. Esto indica que en los últimos años la medida de esta variable ha presentado una tendencia creciente.

**Gráfico 3.** Desocupados que pasan a ser inactivos como proporción del total de desocupados. Total siete ciudades. Enero 2001 - diciembre 2005.



Fuente: ECH-Dane. Cálculos de los autores.

## II. Empalme de series del mercado laboral

Con el propósito de hacer compatibles las series generadas con la ENH (1984 – 2000) y con la ECH (2001 – 2005) para siete ciudades, Suárez y Buriticá (2002) y Lasso (2002) propusieron dos metodologías. Los primeros desarrollaron dos pruebas no paramétricas y una prueba  $t$  para verificar la existencia de valores diferenciados entre la ENH y la ECH. Con esta información construyeron dos factores de corrección: uno para ajustar el cambio en las definiciones de la fuerza de trabajo

y sus componentes y otro para hacer el ajuste por el cambio en el sistema de recolección por trimestre, de cuyo producto resulta un factor definitivo. Cada factor es el resultado de procesos autorregresivos de primer orden en frecuencia trimestral, lo cual permite corregir las estadísticas del mercado laboral del año 2000 hacia atrás. Entre las limitaciones de esta metodología están la imposibilidad de realizar desagregaciones, la escasez de datos con los cuales estimar procesos autorregresivos y la falta de coherencia de los datos obtenidos con el ciclo y la estacionalidad presentados por el mercado laboral (véase Lasso, 2002).

La segunda metodología hace un análisis completo del cambio en la definición de los TFSR y los desocupados y su efecto en las estadísticas laborales de forma separada mediante modelos *logit* con los cuales corrige lo relativo tanto al módulo de fuerza de trabajo (analizado en la sección anterior) como a la estacionalidad. De esta manera, se superan los problemas de la metodología anterior. El trabajo econométrico de Lasso ofrece dos alternativas de empalme, ya que construye dos tipos de modelos para cada caso de reclasificación: uno que usa sólo los meses en los cuales se realizaba la ENH y otro que usa la muestra por trimestres completa con *dummies* mensuales intratrimestre, con el fin de corregir la estacionalidad; así, para cada caso se presentan cinco modelos.

La metodología de Lasso presenta, sin embargo, dos limitaciones. En primer lugar, la inclusión de dos modelos de reclasificación para cada caso no permite tener un resultado único para el empalme. En segundo lugar, aparte la desagregación por género, no se intenta ningún empalme de las tasas por ciudad, grupos de edad, etc.

El presente documento plantea dos alternativas para el empalme de las series del mercado laboral, una de carácter microeconómico y otra basada en la tasa global de participación del mercado. La primera alternativa sigue la metodología de Lasso (2002) al utilizar modelos *logit que* toman como base los datos individuales de la ECH para determinar, a partir de las características personales y familiares, las personas susceptibles de ser reclasificadas. Procura, sin embargo, remediar algunos de los defectos de la propuesta de empalme de dicho autor al incluir elementos que permitan unificar el modelo en uno que contenga

todos los posibles efectos estacionales y una información que genere mayor precisión cuando se estimen altos niveles de desagregación. De igual forma, parte de una teoría para justificar el modelo empírico. La segunda alternativa se desarrolló teniendo en cuenta el hecho empírico surgido de la correlación entre el porcentaje de reclasificados y la TGP, el cual permite determinar el flujo de reclasificados (cuántos) en cada uno de los casos expuestos. Con cada una de las metodologías empleadas para realizar los empalmes se realizan desagregaciones por género y ciudad. A la primera alternativa la llamamos indistintamente microeconómica o *logit*, y a la segunda la llamamos macro, o basada en datos agregados del mercado laboral.

### III. Modelos *logit* para el empalme de series del mercado laboral: enfoque microeconómico

En esta sección se propone un modelo teórico para tratar los ‘nuevos inactivos’ (que vienen de los desocupados que no tienen disponibilidad inmediata para trabajar) y los TFSR (que vienen de los desocupados e inactivos), del cual se derivan dos modelos *logit* cuyos resultados se presentan al final de la misma.

#### A. El tránsito de desocupados a inactivos: enfoque teórico

Se parte del modelo neoclásico de participación según el cual una persona participa en el mercado laboral si el salario de mercado supera su salario de reserva (cuando su nivel *ex-ante* de oferta de horas de trabajo es 0), dadas sus preferencias e ingresos no laborales; es decir, la persona participa si:

$$w^m > w^R \Big|_{l=\bar{l}} \quad \therefore \quad \frac{w^m}{w^R} > 1 \Big|_{l=\bar{l}} \quad (1)$$

ecuación en la cual  $w^m$  es el salario de mercado,  $w^R$  el salario de reserva,  $\bar{l}$  el tiempo total disponible y  $l$  el tiempo destinado al ocio.

El hecho que se quiere caracterizar consiste en estimar la recomposición de la PET y su efecto en las estimaciones de las principales

variables del mercado laboral. Concretamente, se quiere determinar qué ocasiona que personas que en un momento (durante la vigencia de la ENH) fueron consideradas participantes en el mercado laboral pasen a ser consideradas (durante la vigencia de la ECH) parte de la población inactiva. Por tanto, el objetivo es identificar aquellos factores que provocan una caída en salario relativo haciendo que:

$$\frac{w^m}{w^R} < 1 \Big|_{l=\bar{l}} \quad (2)$$

Se considera que el salario de reserva, que es una variable no observable, se puede estimar bajo una especificación que contenga la siguiente forma:

$$w^R = w^R(cp, cf) \quad (3)$$

*cp* son las características de la persona y *cf* las características familiares. En el primer caso, se utilizan variables como edad, escolaridad, nivel educativo, género y *proxies* del ingreso, entre otras. Debe señalarse que este conjunto de características también afecta la oferta salarial que puede recibir el individuo en el mercado. Entre las variables que caracterizan el hogar se tienen en cuenta su localización, los niños menores de seis años, el número de ocupados y desocupados del hogar, entre otras. Existen, desde luego, factores de demanda que afectan el salario de mercado. Sin embargo, aquí sólo nos concentraremos en el lado de la oferta.

## **B. El tránsito de desocupados e inactivos a TFSR: enfoque teórico**

En el caso de los TFSR sólo se presenta un cambio importante respecto al modelo anterior, ya que se puede pensar que es importante considerar variables más específicas de las actividades económicas que realizan los agentes del hogar. Así, suponemos que el salario de reserva depende no sólo de las características personales y familiares, sino también de especificidades del jefe del hogar (*cj*) como la rama de actividad y la posición ocupacional.

$$w^R = w^R(cp, cf, cj) \quad (4)$$

El salario explícito no parece cumplir un papel fundamental en el caso de los TFSR; por tanto, cuando se habla de 'salario' se hace referencia a las ganancias en términos de ayudas en dinero y experiencia y a la satisfacción que obtiene un TFSR por el hecho de prestar ayuda.

### C. Aspectos empíricos

Se estiman dos modelos de elección binaria. Con el primero se quiere determinar la probabilidad de que un individuo que se encuentra como desempleado sea reclasificado como inactivo. Con el segundo, la probabilidad de ser reclasificado como TFSR, si previamente ha sido clasificado como inactivo o desocupado. En cada caso se estima un modelo *logit* de la forma:

$$P(y = 1) = \frac{e^I}{1 + e^I} \quad (5)$$

En esta ecuación  $y$  toma el valor de 1 si el individuo se reclasifica y 0 en otro caso.  $I$  es la función índice que viene dada por una forma lineal:

$$I = \alpha + \sum_{i=1}^9 \left[ \beta_{i0} + \sum_{j=1}^3 \beta_{ij} mt_j \right] H_i + \sum_{i=1}^9 \left[ \phi_{i0} + \sum_{j=1}^3 \phi_{ij} mt_j \right] P_i + \sum_{i=1}^2 \left[ \gamma_{i0} + \sum_{j=1}^3 \gamma_{ij} mt_j \right] J_i + \sum_{i=1}^n \lambda_i T_i + \varepsilon_i \quad (6)$$

siendo  $H$  el vector de características del hogar compuesto de la ciudad de residencia, el estrato socioeconómico, el número de ocupados y desocupados en el hogar, la presencia de niños menores de seis años, la presencia de cónyuge<sup>8</sup> y el género, edad y escolaridad del jefe como

<sup>8</sup> La presencia de niños menores de seis años se modela mediante una *dummy* que toma el valor de 1 si hay al menos un niño menor de seis años. De igual manera se hace para la variable de presencia de cónyuge en el hogar.

conjunto de factores básicos que determinan la participación en el mercado laboral<sup>9</sup>;  $P$ , entre tanto, es el vector de características personales compuesto de la posición en el hogar, el estado civil, el género, la edad, la escolaridad, la asistencia escolar y el nivel escolar más alto alcanzado; el vector  $J$  está compuesto de las ramas de actividad y la posición ocupacional del jefe;  $T$  es un vector que contiene interacciones entre las variables enunciadas anteriormente con las *dummies* de género y ciudad, ya que se quieren analizar estos dos niveles de desagregación;  $mt$  son *dummies* trimestrales y  $\varepsilon$  son innovaciones de carácter aleatorio.

La especificación de este modelo presenta diferencias importantes respecto a los de Lasso (2002). En primer lugar, acá se propone un único modelo que mediante interacciones calcula los efectos estacionales y la estructura serial. En segundo lugar, también mediante interacciones, se capturan los efectos por ciudad y género con el fin de tener información más detallada de estos grupos para cada una de las variables. En último lugar, la especificación del modelo *logit* para los desocupados no incluye las variables de rama de ocupación del jefe del hogar ni la posición ocupacional de éste<sup>10</sup>.

Para efectos empíricos se utilizaron los datos de la ECH en el período 2001-2005 (60 etapas). En el caso de la estimación de la probabilidad de ser reclasificado como inactivo cuando originalmente se clasificó como desocupado, se utilizó una muestra de 114.545 individuos. En el caso de la reclasificación como inactivo o desocupado hacia TFSR, se utilizó una muestra de 471.327 individuos. Con estos datos se construyeron estadísticas<sup>11</sup> con las cuales se da cuenta de la importancia relativa de cada conjunto para determinar la probabilidad que se quiere modelar. Las tablas 1 y 2 del anexo 1 presentan los promedios de cada una de las variables consideradas en el período 2001-2005 y en el período 1984-2000 (correspondiente a 68 etapas de la ENH).

<sup>9</sup> En el modelo de TFSR se incluye la presencia de patrones en el hogar diferentes del jefe de hogar, construida a partir de una variable *dummy* que toma el valor de 1 si hay al menos un individuo en el hogar que tenga la posición de patrón y no sea el jefe.

<sup>10</sup> En el caso de los TFSR se incluyen estas variables pero condicionadas a que el individuo no sea jefe de hogar para evitar la repetición de información. De la misma manera se modelan las variables de género, edad y escolaridad del jefe, tanto para el modelo de desocupados como para el de TFSR.

<sup>11</sup> Expandidas con los factores del Dane.



Con esta información se estiman modelos *logit* para cada caso, haciendo la corrección de Huber-White a los residuales. Con dichos modelos se estima la probabilidad de que un desempleado, de la época de la ENH, fuese clasificado ahora como inactivo y de que un desocupado (inactivo) se clasifique como TFSR con menos de 15 horas. Las tablas 1 y 2 presentan los efectos marginales correspondientes.

Los modelos fueron utilizados para hacer los pronósticos que permitieran reorganizar la información de la fuerza de trabajo corrigiendo en primer lugar la cifra de los desocupados que pasan a ser inactivos y luego la de los desocupados o inactivos que pasan a TFSR. Los resultados de estos ejercicios, incluyendo las estimaciones de Lasso (2002) para modelos trimestrales, aparecen en los gráficos 4, 5 y 6, en los que se presentan las nuevas trayectorias de la TD, la TGP y la TO, respectivamente. En el anexo 2 se presentan las mismas tasas desagregadas por género y ciudad.

**Tabla 1. Modelo *logit* para desocupados susceptibles de ser reclasificados como inactivos. Efectos marginales.**

Variable	Efecto marginal	Variable	Efecto marginal	Variable	Efecto marginal
Bucaramanga	-0,05825	Hombre × Trim I	0,01233	Cali × escolaridad	0,00752
Barranquilla	-0,07727	Edad × Trim II	-0,00247	Cali × escolaridad cuadrado	-0,00031
Cali	-0,02757	Edad cuadrado × trim II	0,00003	Manizales × escolaridad	0,01664
Medellín	0,05465	Estudia actualmente × trim I	0,01525	Manizales × esc. cuadrado	-0,00076
Manizales	-0,00164 *	Estudia actualmente × trim II	0,01733	Pasto × escolaridad	0,01764
Pasto	-0,02979 *	Estudia actualmente × trim III	0,01404	Pasto × escolaridad cuadrado	-0,00079
Estrato bajo bajo	-0,01871	Desocupados hogar × trim III	0,00573	Bucaramanga × estudia act.	0,02512
Estrato bajo	-0,01688	Bucaramanga × estrato medio	0,02267	Cali × estudia act.	-0,02776
Estrato medio bajo	-0,01275	Medellín × estrato bajo bajo	-0,02407	Bucaramanga × primaria comp.	-0,01564
Estrato medio	-0,01642	Medellín × estrato bajo	-0,02103	Bucaramanga × secund. inc.	-0,02009
Jefe	0,06919	Medellín × estrato medio bajo	-0,01838	Barranquilla × primaria incomp.	0,05863
Cónyuge	0,06210	Manizales × estrato bajo bajo	0,02780	Cali × primaria completa	-0,02558
Hijo	0,03469	Pasto × estrato bajo	-0,03846	Cali × secundaria incompleta	-0,02451
Otro pariente	0,03001	Pasto × estrato medio bajo	-0,02108	Cali × secundaria completa	-0,02246
Unión libre	0,03934	Barranquilla × hijo	-0,03535	Medellín × primaria incompleta	0,02014
Casado	0,04227	Barranquilla × otro pariente	-0,01915	Manizales × primaria completa	-0,03690
Viudo, separado o divorciado	0,01217	Cali × jefe	-0,02458	Manizales × secundaria inc.	-0,04021
Género hombre	-0,02172 *	Cali × cónyuge	-0,01891	Manizales × secundaria comp.	-0,03660
Edad	-0,00862	Cali × hijo	-0,01297	Pasto × primaria completa	-0,04244
Edad cuadrado	0,00012	Medellín × jefe	-0,04920	Pasto × secundaria incompleta	-0,04633

(Continúa...)

Tabla 1. Modelo *logit* para desocupados susceptibles de ser reclasificados como inactivos. Efectos marginales. (...Continuación).

Variable	Efecto marginal	Variable	Efecto marginal	Variable	Efecto marginal
Estudia actualmente	0,13867	Medellín × cónyuge	-0,03617	Pasto × secundaria completa	-0,02756
Sin educación	0,09796	Medellín × hijo	-0,02937	Manizales × jefe hombre	0,01148 *
Primaria incompleta	0,04041	Medellín × otro pariente	-0,02638	Manizales × edad jefe	-0,00029
Primaria completa	0,06913	Manizales × jefe	-0,03915	Medellín × escolaridad jefe	-0,00116
Secundaria incompleta	0,05619	Manizales × cónyuge	-0,04320	Pasto × escolaridad jefe	-0,00184
Secundaria completa	0,03059	Manizales × hijo	-0,03250	Bucaramanga × ocupados hog.	-0,00415
Ocupados en el hogar	0,00350	Manizales × otro pariente	-0,02751	Barranquilla × ocupados hog.	-0,00775
Desocupados en el hogar	-0,01665	Pasto × jefe	-0,05744	Barranquilla × desocupados hog.	0,00836
Trimestre I	-0,01776	Pasto × cónyuge	-0,05104	Cali × desocupados hogar	0,00562
Trimestre II	0,03819	Pasto × hijo	-0,05862	Medellín × desocupados hogar	0,00539
Trimestre III	-0,00766	Pasto × otro pariente	-0,05039	Pasto × desocupados hogar	0,00688
Jefe hombre	-0,00462 *	Barranquilla × unión libre	0,02819	Hombre × hijo	-0,01870
Edad del jefe	-0,00010 *	Barranquilla × casado	0,02507 *	Hombre × otro pariente	-0,01381 *
Escolaridad del jefe	0,00142	Barranquilla × viudo, sep. o div.	0,02553 *	Hombre × unión libre	-0,04919
Barranquilla × trim I	0,03073	Medellín × unión libre	0,02719	Hombre × casado	-0,04048
Barranquilla × trim II	0,06860	Medellín × casado	0,02688	Hombre × viudo, sep. o div.	-0,04197
Barranquilla × trim III	0,03622	Medellín × viudo, sep. o div.	0,01653	Hombre × edad	-0,00222
Cali × trim I	0,01449	Pasto × viudo, sep. o div.	-0,02301	Hombre × edad cuadrado	0,00003
Cali × trim II	-0,01616	Barranquilla × hombre	-0,01676	Hombre × escolaridad	0,00139
Medellín × trim III	-0,01294	Manizales × hombre	-0,01191	Hombre × estudia actualmente	0,04392
Pasto × trim I	0,08967	Bucaramanga × edad	0,00185	Hombre × secundaria completa	0,01714
Pasto × trim II	0,02072	Bucaramanga × edad cuadrado	-0,00002 *	Hombre × jefe hombre	0,01534
Pasto × trim III	0,04839	Barranquilla × escolaridad	0,00877	Hombre × desocupados hogar	0,00795
Hijo × trim II	0,01070	Barranquilla × esc. cuadrado	-0,00038	Constante	--
<b>Observaciones</b>	114.545	<b>Concordancia</b>	0,76	<b>Pseudo R<sup>2</sup></b>	0,13

\* Coeficiente no significativo al 10%.

Fuente: cálculos de los autores.

Si bien los resultados permiten realizar una reclasificación confiable, la concordancia y el ajuste sugieren buscar otra alternativa de empalme. Así, teniendo en cuenta los agregados del mercado laboral se encuentran estrechas correlaciones que permiten estimar la recompo-

sición de la población económicamente activa y la población en edad de trabajar entre 1984 y 2000. Dichos resultados se presentan en la siguiente sección.

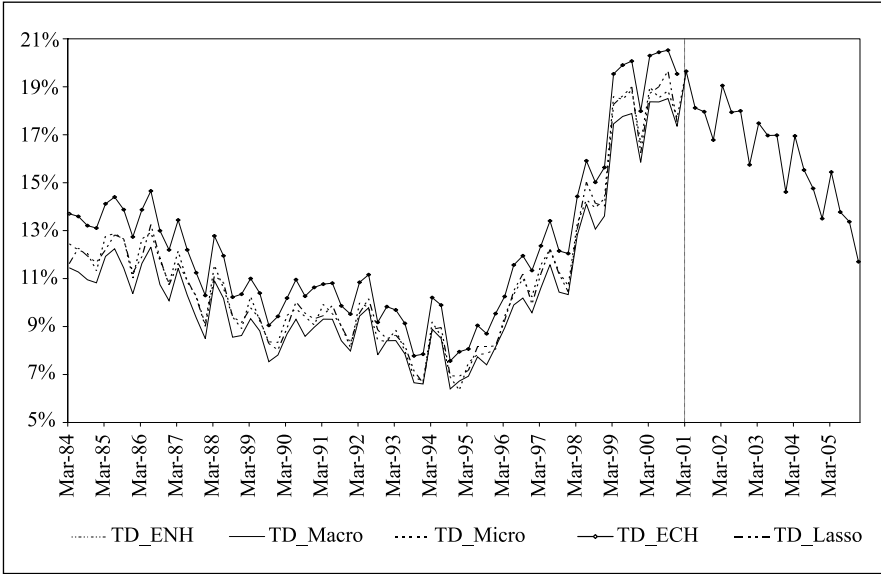
**Tabla 2. Modelo logit para desocupados e inactivos susceptibles de ser reclasificados como TFSR. Efectos marginales.**

Variable	Efecto marginal	Variable	Efecto marginal	Variable	Efecto marginal
Bucaramanga	0,00091	Jefe patrón	0,00929	Pasto x estudia act.	0,00090
Barranquilla	0,00406 *	Patrón diferente del jefe	0,00283	Barranquilla x jefe hombre	0,00204
Cali	0,01332	Bucaramanga x trim II	0,00179	Cali x jefe hombre	0,00076
Medellín	0,00350 *	Cali x trim II	0,00127	Medellín x jefe hombre	0,00070
Manizales	0,01867 *	Medellín x trim III	0,00088	Medellín x edad jefe	0,00003
Pasto	0,00067	Manizales x trim II	0,00098	Cali x jefe sector construcción	0,00311
Estrato bajo bajo	0,00341	Pasto x trim I	0,00109	Cali x jefe sector comercio	0,00058
Estrato bajo	0,00234	Pasto x trim II	0,00104	Cali x jefe sector transp.	0,00230
Estrato medio bajo	0,00176	Jefe x trim I	-0,00163	Cali x jefe sector serv. sociales	0,00238
Estrato medio	0,00167	Cónyuge x trim I	-0,00147	Medellín x jefe sector agrop.	0,00318
Jefe	0,01113	Hijo x trim I	-0,00148	Medellín x jefe sector construc.	0,00196
Cónyuge	0,02949	No pariente x trim I	-0,00143	Manizales x jefe sector agrop.	0,00181
Hijo	0,01563	Hombre x trim II	-0,00055	Manizales x jefe sector finan.	0,00568
Otro pariente	0,02209	Manizales x estrato bajo	-0,00058	Pasto x jefe sector construcción	0,00440
Unión libre	-0,00056	Manizales x estrato medio	-0,00067	Pasto x jefe sector transp.	0,00276
Casado	0,00019 *	Pasto x estrato bajo	-0,00130	Pasto x jefe sector serv. sociales	0,00270
Viudo, separado o divorciado	-0,00114	Edad x trim II	-0,00004	Bucaramanga x jefe TFSR	-0,00155
Género hombre	0,00220	Edad cuadrado x trim II	0,0000003	Bucaramanga x jefe cuenta p.	-0,00072
Edad	0,00007	Bucaramanga x hijo	-0,00095	Barranquilla x jefe empleado	-0,00191
Edad cuadrado	-0,000001	Bucaramanga x otro pariente	-0,00106	Barranquilla x jefe cuenta p.	-0,00161
Estudia actualmente	0,00057	Barranquilla x jefe	-0,00242	Barranquilla x jefe patrón	-0,00140
Escolaridad	0,00024	Barranquilla x cónyuge	-0,00260	Cali x jefe TFSR	-0,00173
Escolaridad cuadrado	-0,00001	Barranquilla x hijo	-0,00256	Cali x jefe cuenta propia	-0,00098
Ocupados en el hogar	0,00066	Barranquilla x otro pariente	-0,00231	Cali x jefe patrón	-0,00087
Desocupados en el hogar	-0,00115	Cali x jefe	-0,00196	Medellín x jefe empleado	0,00194
Trimestre I	0,00180 *	Cali x cónyuge	-0,00215	Medellín x jefe patrón	0,00118
Trimestre II	-0,00056 *	Cali x hijo	-0,00219	Pasto x jefe empleado	-0,00169
Trimestre III	-0,00060	Cali x otro pariente	-0,00217	Pasto x jefe cuenta propia	-0,00115
Jefe hombre	-0,00096	Medellín x jefe	-0,00178	Pasto x jefe patrón	-0,00136
Edad jefe	-0,00002	Medellín x cónyuge	-0,00248	Bucaramanga x patrón no jefe	-0,00090
Escolaridad jefe	-0,00006	Medellín x hijo	-0,00267	Medellín x patrón no jefe	-0,00109
Jefe sector agropecuario	-0,00020 *	Medellín x otro pariente	-0,00232	Hombre x edad	-0,00007
Jefe sector industrial	0,00278	Manizales x jefe	-0,00188	Hombre x edad cuadrado	0,00000
Jefe sector construcción	-0,00170	Manizales x cónyuge	-0,00185	Hombre x jefe hombre	0,00090
Jefe sector comercio	0,00283	Manizales x hijo	-0,00188	Hombre x edad Jefe	-0,00003
Jefe sector transp. y comunic.	-0,00168	Manizales x otro pariente	-0,00194	Hombre x ocupados hogar	0,00029
Jefe sector financiero	-0,00079 *	Bucaramanga x viudo, sep.	0,00192	Hombre x jefe sector transp.	0,00189
Jefe sector servicios sociales	-0,00086	Pasto x hombre	-0,00044	Hombre x jefe TFSR	-0,00110
Jefe TFSR	0,06692	Manizales x escolaridad	-0,00023	Hombre x jefe cuenta propia	-0,00040
Jefe empleado	-0,00104	Manizales x escolaridad cuad.	0,00001	Constante	-
Jefe cuenta propia	0,00681	Bucaramanga x estudia act.	0,00167		
<b>Observaciones</b>	<b>471.327</b>	<b>Concordancia</b>	<b>0,876</b>	<b>Pseudo R<sup>2</sup></b>	<b>0,191</b>

\* Coeficiente no significativo al 10%.

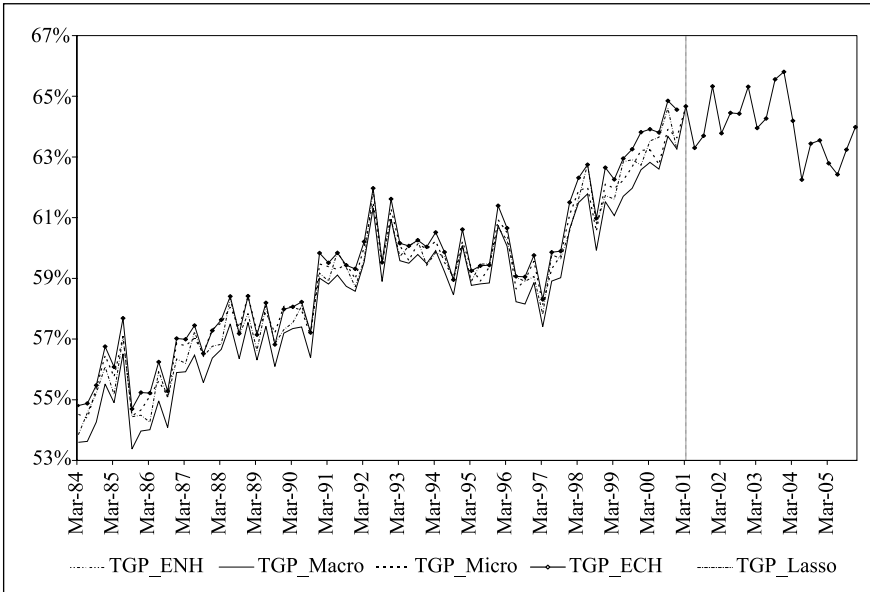
Fuente: cálculos de los autores.

**Gráfico 4. Tasa de desempleo siete ciudades con metodología de ENH y ECH. 1984 - 2005.**



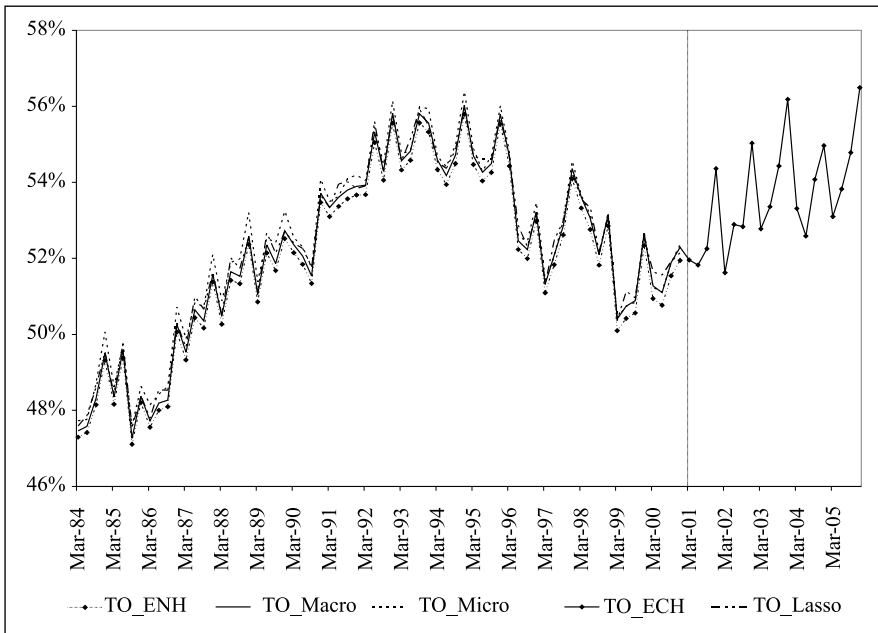
Fuente: ENH 1984-2000 y ECH 2001-2005. Cálculos de los autores.

**Gráfico 5. Tasa global de participación siete ciudades con metodología de ENH y ECH. 1984 - 2005.**



Fuente: ENH 1984-2000 y ECH 2001-2005. Cálculos de los autores.

Gráfico 6. Tasa de ocupación siete ciudades con metodología de ENH y ECH. 1984 - 2005.



Fuente: ENH 1984-2000 y ECH 2001-2005. Cálculos de los autores.

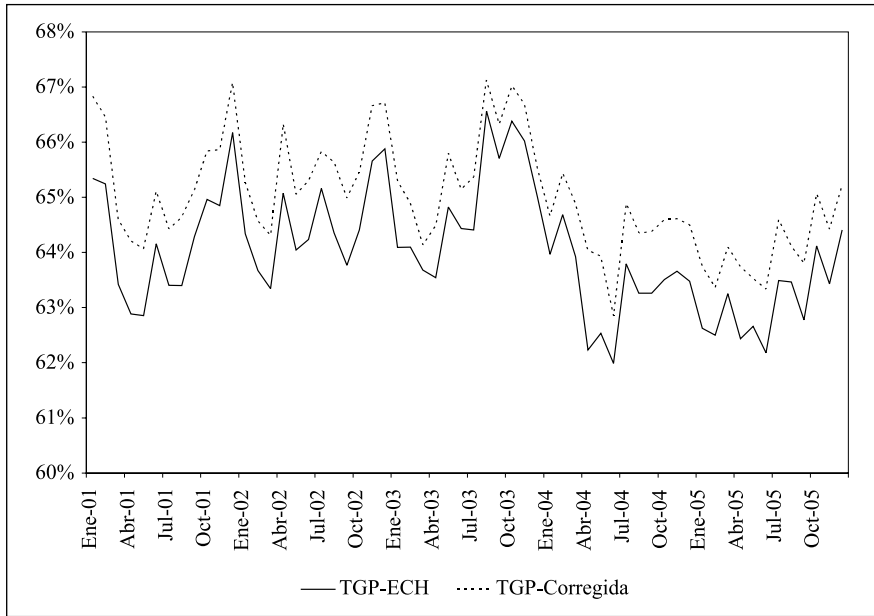
#### IV. Modelos basados en variables agregadas del mercado laboral: enfoque agregado ('macro')

Después de analizar las dinámicas de la TD, la TGP y la TO, se encontró una estrecha relación entre el comportamiento de la TGP, siete ciudades, y el porcentaje de reclasificados tanto para los desocupados como para los TFSR; en el primer caso, se observa una relación negativa que indica que un aumento en la TGP implica la disminución del porcentaje de desocupados que no tienen disponibilidad inmediata para trabajar, mientras que para el segundo caso el signo de la correlación cambia.

Se encontró, entonces, que la TGP es un buen mecanismo para pronosticar el porcentaje de reclasificados en los casos ya mencionados, pero dado que el objetivo es predecir dicho comportamiento en el período 1984-2000, la TGP debía ser corregida bajo la metodología de ENH (véase gráfico 7). Como resultado, la TGP, corregida con la

metodología de ENH, es, en promedio, 0,98 *pp* más alta que la misma tasa construida con la metodología de la ECH. De todas maneras, al corregir la TGP, se mantuvo la correlación existente con los flujos de poblaciones reclasificadas.

Gráfico 7. TGP ECH y TGP corregida a ENH. Siete ciudades. Enero 2001 - diciembre 2005.



Fuente: ENH y ECH. Cálculos de los autores.

Para la corrección de la TGP se identificaron los desocupados reclasificados como inactivos, y los inactivos y desocupados reclasificados como TFSR. Para ello, se clasificaron como desocupados o inactivos aquellos TFSR que trabajaban entre 1 y 14 horas y como desocupados, aquellos individuos que cumplían con las características de los desocupados, excepto por la disponibilidad inmediata a trabajar o tener una razón válida de desempleo.

En el caso del flujo de inactivos-desocupados, no se presentan problemas de identificación, pero en el caso de los TFSR no se tiene suficiente información para conocer si estas personas fueron consideradas inactivas o desocupadas en la ENH. Para resolver este inconveniente se adoptó el criterio de hacer el ajuste de tal forma que las proporciones

de desocupados e inactivos respecto del total de ellos se mantuvieran. Esto es, que:

$$\frac{D}{D + PEI} = \frac{D'}{D' + PEI'}, \text{ y } \frac{PEI}{D + PEI} = \frac{PEI'}{D' + PEI'} \quad (7)$$

En la cual  $D$  y  $PEI$  son los desocupados e inactivos en ECH, respectivamente, y  $D'$  y  $PEI'$  sus similares para la ENH.

A partir de estos datos se estimaron modelos OLS para el período 2001-2005, en frecuencia mensual, cuyas variables dependientes correspondían a los porcentajes de reclasificados en cada caso expuesto y utilizaban la *TGP corregida* de la ECH como única variable explicativa, que controla por efectos estacionales en frecuencia trimestral.

Empíricamente se estimó una ecuación de la forma:

$$y_i = \alpha + \sum_{j=1}^3 \beta_j dt_{ji} + \delta TGP \text{ corregida}_i + \sum_{j=1}^T \lambda_j ddj_i + \varepsilon_i \quad (8)$$

en la cual  $y$  es el porcentaje de reclasificados en cada caso (véanse gráficos 2 y 3),  $dt_i$  *dummies* de frecuencia trimestral, *TGP corregida* la tasa global de participación corregida con metodología de ENH, *ddj* *dummies* de corrección para *outliers* y cambios estructurales y  $\varepsilon_i$  innovaciones. Una vez ajustada la ecuación de regresión (haciendo la corrección de Newey-West para heterocedasticidad y correlación serial) y las pruebas de estabilidad de parámetros de ambos modelos<sup>12</sup>, se obtuvieron los resultados presentados en las tablas 3 y 4 para el caso de los desocupados y los TFSR, respectivamente.

Se observa allí que los signos de los coeficientes correspondientes a la *TGP corregida* son los esperados, dadas las correlaciones negativa y positiva que tienen los flujos de dichas poblaciones reclasificadas. En el anexo 1 se verifica que en períodos en los cuales la TGP era relativamente más alta, el porcentaje de reclasificados era menor en el

<sup>12</sup> Las pruebas de estabilidad son: residuales recursivos, CUSUM, CUSUM cuadrado y coeficientes recursivos.

caso de los desocupados y mayor para el caso de los TFSR con menos de 15 horas. A partir de los resultados de estos modelos se pronosticó el porcentaje de reclasificados durante la vigencia de la ENH y con esta información se corrigieron las tasas de desempleo considerando en primer lugar el efecto del paso de desocupados a inactivos y de desocupados e inactivos a TFSR. Estos resultados se presentan en los gráficos 4, 5 y 6.

**Tabla 3. Modelo para desocupados reclasificados como inactivos.**

Variable	Coefficiente	Estadístico t	Valor-p
<i>TGP corregida</i>	-0,8383	-3,3935	0,0013
Trimestre I	-0,0249	-4,0851	0,0002
Trimestre II	-0,0178	-2,7439	0,0084
Trimestre III	-0,0133	-2,5353	0,0143
<i>dummy I</i>	0,0481	12,6134	0,0000
<i>dummy II</i>	-0,0449	-9,6570	0,0000
<i>dummy III</i>	0,0352	5,6164	0,0000
<i>dummy IV</i>	0,0152	2,9554	0,0047
Constante	0,6650	4,0567	0,0002
R <sup>2</sup>	0,6390	Método: mínimos cuadrados ordinarios	
Estadístico F	11,2872	Muestra: 2001,1 – 2005,12	
Valor-p (Estadístico F)	0,0000	Observaciones: 60	
Jarque –Bera	0,92359	Valor-p (J-B) = 0,630152	

*Fuente:* cálculos de los autores.

**Tabla 4. Modelo para desocupados e inactivos reclasificados como TFSR.**

Variable	Coefficiente	Estadístico t	Valor-p
<i>TGP corregida</i>	0,0415	2,0990	0,0403
<i>dummy I</i>	0,0046	6,9715	0,0000
Constante	-0,0197	-1,5269	0,1323
R <sup>2</sup>	0,5408	Método: mínimos cuadrados ordinarios	
Estadístico F	33,5739	Muestra: 2001,1 – 2005,12	
Valor-p (Estadístico F)	0,0000	Observaciones: 60	
Jarque –Bera	0,75951	Valor-p (J-B) = 0,68403	

*Fuente:* cálculos de los autores.



Utilizando este mismo enfoque se ajustaron los agregados del mercado laboral por género y por ciudades; los resultados se resumen en los anexos 2 y 3. Para la desagregación por género se usa la misma metodología que para el modelo agregado del total de siete ciudades, es decir, se estiman modelos OLS haciendo la corrección de Newey-West (véase tabla 5).

Para el caso de la desagregación por ciudad se estimó un *panel* por tipo de reclasificación ya que los modelos individuales no reportaron resultados satisfactorios. Los coeficientes obtenidos por el modelo *panel* parecen ser compatibles con los obtenidos en los modelos individuales, por lo menos en magnitud. El *panel* se estimó por efectos aleatorios teniendo en cuenta los resultados de la prueba de Hausman. Los resultados aparecen en la tabla 6.

El cambio de metodología provocó una reducción en la medida de la tasa de desempleo, cuyo efecto se concentra básicamente en la reclasificación de los desocupados como inactivos, pues este factor contribuye casi con 95% del cambio en el nivel de la tasa de desempleo. Para los indicadores agregados de siete ciudades, los modelos construidos a partir la *TGP corregida* arrojan un mayor cambio en comparación con los modelos *logit*: son en promedio de 1,7 y 1,2 *pp* en el caso de la TD, y de 0,9 y 0,4 *pp* en la TGP y aumentos de 0,2 y 0,5 *pp* en la TO, respectivamente.

Tabla 5. Modelos para desocupados reclasificados como inactivos. Desagregación por género.

Modelo	<i>TGP corregida</i>	R <sup>2</sup> ajustado	Prueba de contraste global		Prueba de normalidad residual	
			Estadístico F	Valor-p	Jarque Bera	Valor-p
Hombres	-0,7837	0,5185	13,7115	0,0000	1,2224	0,5426
Mujeres	-0,4879	0,4893	7,2827	0,0000	0,9120	0,6337
Modelos agregados para el caso de los TFSR.						
Modelo	<i>TGP corregida</i>	R <sup>2</sup> ajustado	Prueba de contraste global		Prueba de normalidad residual	
			Estadístico F	Valor-p	Jarque Bera	Valor-p
Hombres	-0,0438*	0,7557	21,2788	0,0000	0,9722	0,6150
Mujeres	-0,0168*	0,5281	8,3391	0,0000	0,7376	0,6915

\* Coeficiente no significativo al 10%.

Fuente: cálculos de los autores.

Tabla 6. Modelo panel para ciudades.

Modelo desocupados		Modelo TFSR	
<i>TGP corregida</i>	0,4658	<i>TGP corregida</i>	0,0741
E.E. <i>TGP corregida</i>	0,0705	E.E. <i>TGP corregida</i>	0,0092
$\sigma_e$	0,0315	$\sigma_e$	0,0045
$\sigma_u$	0,0386	$\sigma_u$	0,0020
$\rho$	0,6007	P	0,1620
$R^2$ total	0,1245	$R^2$ total	0,3273

Fuente: cálculos de los autores.

Las estimaciones por género presentan resultados similares a los datos agregados, pero las tasas empalmadas con la metodología macro sufren variaciones de hasta 1 *pp* en comparación con los resultados obtenidos con el modelo *logit*, como ocurre en el caso de la tasa de desempleo para las mujeres. Sin embargo, a mayor nivel de desagregación, como es el caso por ciudades, los modelos elaborados con información individual tienen mayor capacidad que los macro; pues, a este nivel de desagregación, la correlación existente entre los porcentajes de reclasificados y la TGP corregida para cada ciudad es muy baja, en especial para el caso de los modelos de los TFSR. Así, cuanto mayor es el nivel de desagregación (por género o ciudad) mayor es el poder de predicción del modelo microeconómico (*logit*) en cuanto a la dinámica del mercado laboral.

Lo anterior señala la pertinencia de definir cuál de las estimaciones es más apropiada para realizar análisis que abarquen el período 1984-2005: la *logit* o la macro. El primer criterio es que la serie mantenga las propiedades estadísticas de la serie original. El segundo es la capacidad de pronosticar tanto para el año 2000, utilizando la muestra 1984-1999, como para el período 2001 y 2001-2002 utilizando la muestra completa (1984-2000).

La tabla 3 del anexo 4 presenta las pruebas de raíz unitaria (ADF y ADF-ERS, en este caso) tanto para la serie original como para las estimadas bajo los dos métodos. Se puede observar que las propiedades seriales de las estimaciones por ambos modelos se mantienen; además, se presentan el segundo, tercer y cuarto momentos ‘muestrales’ de la primera diferencia tanto de la serie de ENH como de las series estimadas y se observa, en general, bastante similitud entre éstas, especialmente entre la serie generada por el modelo microeconómico (*logit*) y la serie de ENH. A pesar de esto, es difícil afirmar que la

serie del modelo micro es la mejor, ya que los cambios producidos en la encuesta de hogares pudieron haber variado la estructura de la serie, pues como se demostró en el modelo agregado, estos cambios tienen una estrecha relación con la TGP, la cual ha sufrido variaciones importantes en el período 1984-2000.

Con respecto al segundo criterio, se empleó la raíz del error cuadrático medio de pronóstico, comparando los valores obtenidos, para el 2000, con cada serie estimada y los valores oficiales de ECH para el mismo año; y, además, se estimó un modelo ARIMA para las series obtenidas por ambas metodologías (el mismo modelo ARIMA se aplicó a las dos series estimadas), para realizar pronósticos a un horizonte de ocho trimestres, es decir, 2001:1-2002:4<sup>13</sup>, para así definir cuál de las series tenía mayor precisión para explicar la dinámica del mercado laboral (véase tabla 4 del anexo 4).

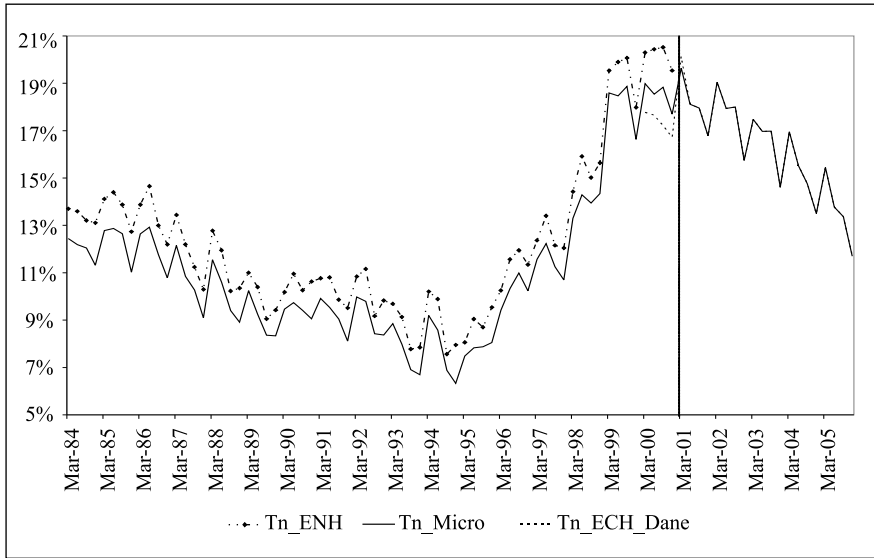
Con base en este análisis comparativo se puede señalar que para la tasa de desempleo urbano (siete ciudades) y para la correspondiente a cada una de las siete ciudades es más adecuado usar estimaciones basadas en el modelo microeconómico (*logit*). Sin embargo, por género la decisión cambia, aunque la diferencia entre los modelos es bastante pequeña<sup>14</sup>. En tal sentido, identificamos las estimaciones del modelo microeconómico como las más adecuadas para realizar el empalme con las series del mercado laboral derivadas de la ECH.

Una vez sugerida una serie, vale la pena señalar que los datos presentados en este documento tienen una diferencia con los datos oficiales presentados por el Dane para el período comprendido entre el primer trimestre de 2000 y el primer trimestre de 2001. Tal diferencia se debe a que, como lo advierte el pie de página 1, en el año 2000 todavía no se habían realizado todos los cambios metodológicos y éstos terminaron solamente en junio de 2001. Por tal razón, para ese período también se realizaron estimaciones al igual que se hizo para el resto del período de la ENH. Las estimaciones nuestras de las tasas de desempleo, participación y ocupación con las correspondientes series oficiales del Dane se presentan en los gráficos 8 a 10. Las diferencias aparecen entonces durante el año 2000 y el primer semestre de 2001.

<sup>13</sup> Se estimaron varios modelos ARIMA y se utilizaron los criterios de información de Schwarz y Akaike para la elección del mejor modelo.

<sup>14</sup> Para la desagregación por género, no se tienen datos oficiales para comparar el año 2000.

**Gráfico 8.** Tasas de desempleo oficiales del Dane y tasa estimada con el modelo microeconómico (micro). Siete ciudades. 1984 - 2005.



Fuente: ECH y ENH. Cálculos de los autores.

**Gráfico 9.** Tasas de globales de participación oficiales del Dane y estimada con el modelo microeconómico (micro). Siete ciudades. 1984 - 2005.

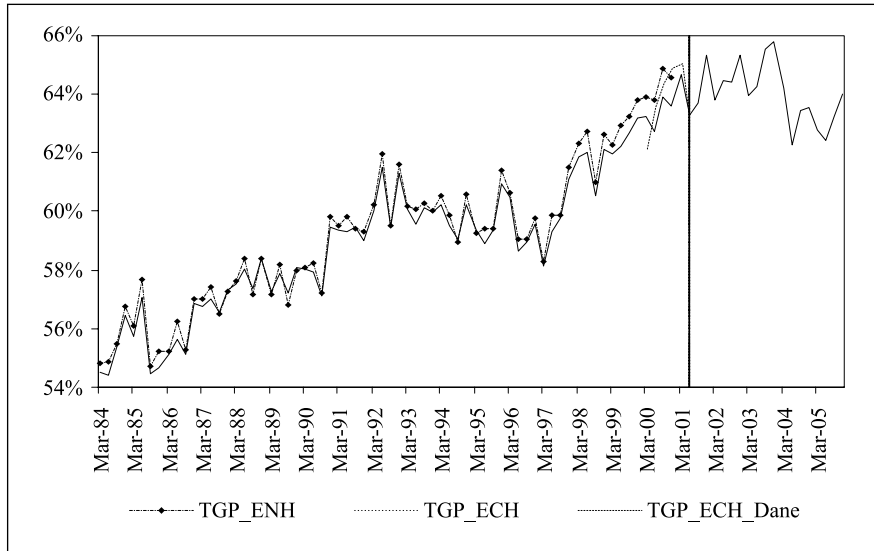
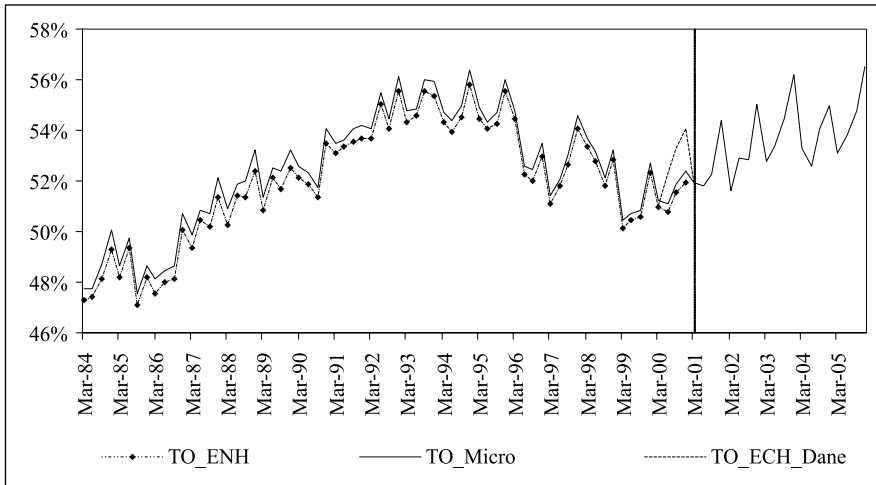


Gráfico 10. Tasas de ocupación oficiales del Dane y estimada con el modelo microeconómico (micro). Siete ciudades. 1984 - 2005.



## V. Conclusiones

El sistema de encuestas de hogares sufrió un profundo cambio metodológico, lo cual ha permitido tener un flujo de información más frecuente y nutrido, pero esto causó una discontinuidad en las series del mercado laboral. Esto ha hecho más difícil realizar análisis de largo plazo y ha obligado a proponer alternativas para compatibilizar la vieja metodología (ENH) con la actual (ECH).

El criterio de disponibilidad inmediata para trabajar, en el caso de la medición del desempleo, ha causado el mayor efecto en las series del mercado laboral, pues entre 2001 y 2005 se reclasificaron en promedio 11% de los desocupados, lo cual redujo la medida de la tasa de desempleo hasta en 1,9 *pp*. Mientras que la reclasificación de los TFSR con menos de 15 horas, que en promedio se reclasificaron 0,8% entre desocupados e inactivos, tan sólo alcanza a reducir la tasa de desempleo en 0,1 *pp*.

Con el propósito de usar los datos del mercado laboral desde 1984 hasta ahora, presentamos dos metodologías de empalme. La primera sigue el enfoque de Lasso (2002), mientras que la segunda se basa en algunas regularidades observadas en el caso del mercado laboral de Colombia.

El enfoque basado en variables agregadas presenta buenos resultados en niveles de alta agregación, mientras que en el caso contrario 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. La base de datos con las series se encuentra disponible, junto al *Borrador de Economía*, en la página del Banco de la República en Internet.

## Referencias

- DANE. (2001). *Metodología de la Encuesta Continua de Hogares*.
- GUATAQUÍ, J. C. and TABORDA, R. (2006). “Theoretical and empirical implications of the new definition of unemployment in Colombia”, *Revista de Economía del Rosario*, 9(1):21-38 .
- LASSO, F. (2002). “Nueva metodología de encuesta de hogares: ¿más o menos desocupados?”, *Archivos de Economía*, 213, DNP.
- SUÁREZ, Á. y BURITICÁ, A. (2002). “Empalme de las series de tasa de desempleo, ocupación y participación entre la Encuesta Transversal y la Continua”, Dane.

## Anexo 1

Tabla 1. Promedio de las principales variables del modelo *logit* para el caso de los desocupados.

y	ENH		ECH	
	Total	Total	0	1
<b>Frecuencia</b>	<b>280.304</b>	<b>115.502</b>	<b>105.472</b>	<b>10.030</b>
Bucaramanga	0,0628	0,0684	0,0713	0,0445
Barranquilla	0,0940	0,0799	0,0869	0,0229
Bogotá	0,3969	0,4721	0,4563	0,6010
Cali	0,1648	0,1470	0,1497	0,1251
Medellín	0,2270	0,1821	0,1833	0,1718
Manizales	0,0277	0,0265	0,0266	0,0252
Pasto	0,0268	0,0241	0,0259	0,0095
Estrato bajo bajo	0,0726	0,0695	0,0716	0,0524
Estrato bajo	0,2849	0,2965	0,2996	0,2715
Estrato medio bajo	0,4776	0,4435	0,4424	0,4525
Estrato medio	0,1256	0,1293	0,1275	0,1437
Estrato medio alto y alto	0,0392	0,0611	0,0588	0,0800
Jefe	0,1675	0,2050	0,2129	0,1404
Cónyuge	0,1796	0,1933	0,1854	0,2575
Hijo	0,5068	0,4521	0,4497	0,4718
Pariente	0,1284	0,1324	0,1344	0,1165
No pariente	0,0145	0,0169	0,0174	0,0135
Unión libre	0,1459	0,1940	0,1936	0,1972
Casado	0,1946	0,1842	0,1853	0,1752
Viudo, separado o divorciado	0,0910	0,1007	0,1040	0,0735
Soltero	0,5684	0,5211	0,5171	0,5541
Hombre	0,4426	0,4402	0,4520	0,3444
Mujer	0,5574	0,5598	0,5480	0,6556
Edad	28,1537	30,6771	30,9321	28,6022
Escolaridad	9,6621	11,3861	11,3870	11,3787
Asistencia escolar	0,1509	0,1583	0,1305	0,3846
Sin educación	0,0153	0,0112	0,0109	0,0132
Primaria incompleta	0,0952	0,0610	0,0626	0,0485
Primaria completa	0,1420	0,1094	0,1109	0,0969
Secundaria incompleta	0,3415	0,2450	0,2363	0,3161

(Continúa...)

Tabla 1. Promedio de las principales variables del modelo *logit* para el caso de los desocupados. (...Continuación).

y	ENH		ECH	
	Total	Total	0	1
Educación superior	0,1434	0,2637	0,2585	0,3056
Jefe de hogar hombre	0,7298	0,6822	0,6798	0,7012
Jefe de hogar mujer	0,2702	0,3178	0,3202	0,2988
Edad jefe	47,2116	48,0965	48,2704	46,6810
Escolaridad jefe	7,7555	9,4529	9,3513	10,2819
Hogar sin cónyuge	0,3090	0,3382	0,3413	0,3131
Ocupados en el hogar	1,5113	1,3543	1,3464	1,4183
Desocupados en el hogar	0,5580	0,7310	0,7487	0,5867
Niños menores de 6 años	0,2201	0,1747	0,1725	0,1928
TGPc	0,6009	0,6493	0,6494	0,6487

Nota: los valores presentados se leen como la participación sobre la muestra total, excepto para el caso de la edad, la escolaridad, los ocupados y desocupados en el hogar. Es decir, en el total de la muestra en la ECH para el modelo de los desocupados, se tiene 6,84% de personas que se encuentran en Bucaramanga, considerando los individuos no reclasificados ( $y = 0$ ) un 7,13% y entre los reclasificados como inactivos ( $y = 1$ ) se tiene que 4,45% de individuos de esta ciudad.

Fuente: ECH y ENH. Cálculos de los autores.

Tabla 2. Promedio de las principales variables del modelo *logit* para el caso de los TFSR.

y	ENH		ECH	
	Total	Total	0	1
<b>Frecuencia</b>	<b>1.828.587</b>	<b>475.792</b>	<b>471.842</b>	<b>3.950</b>
Bucaramanga	0,0591	0,0601	0,0599	0,0842
Barranquilla	0,1140	0,1159	0,1167	0,0195
Bogotá	0,4081	0,4289	0,4289	0,4313
Cali	0,1448	0,1432	0,1425	0,2306
Medellín	0,2212	0,2022	0,2025	0,1712
Manizales	0,0307	0,0273	0,0273	0,0310
Pasto	0,0220	0,0223	0,0223	0,0322
Estrato bajo bajo	0,0590	0,0648	0,0647	0,0718
Estrato bajo	0,2432	0,2623	0,2620	0,2979
Estrato medio bajo	0,4691	0,4340	0,4339	0,4516
Estrato medio	0,1551	0,1425	0,1426	0,1303
Estrato medio alto y alto	0,0736	0,0964	0,0968	0,0484
Jefe	0,1439	0,1994	0,2006	0,0453
Cónyuge	0,2712	0,2213	0,2215	0,1947
Hijo	0,4342	0,4098	0,4078	0,6665
Pariente	0,1361	0,1546	0,1551	0,0882
No pariente	0,0106	0,0141	0,0142	0,0051

(Continúa...)



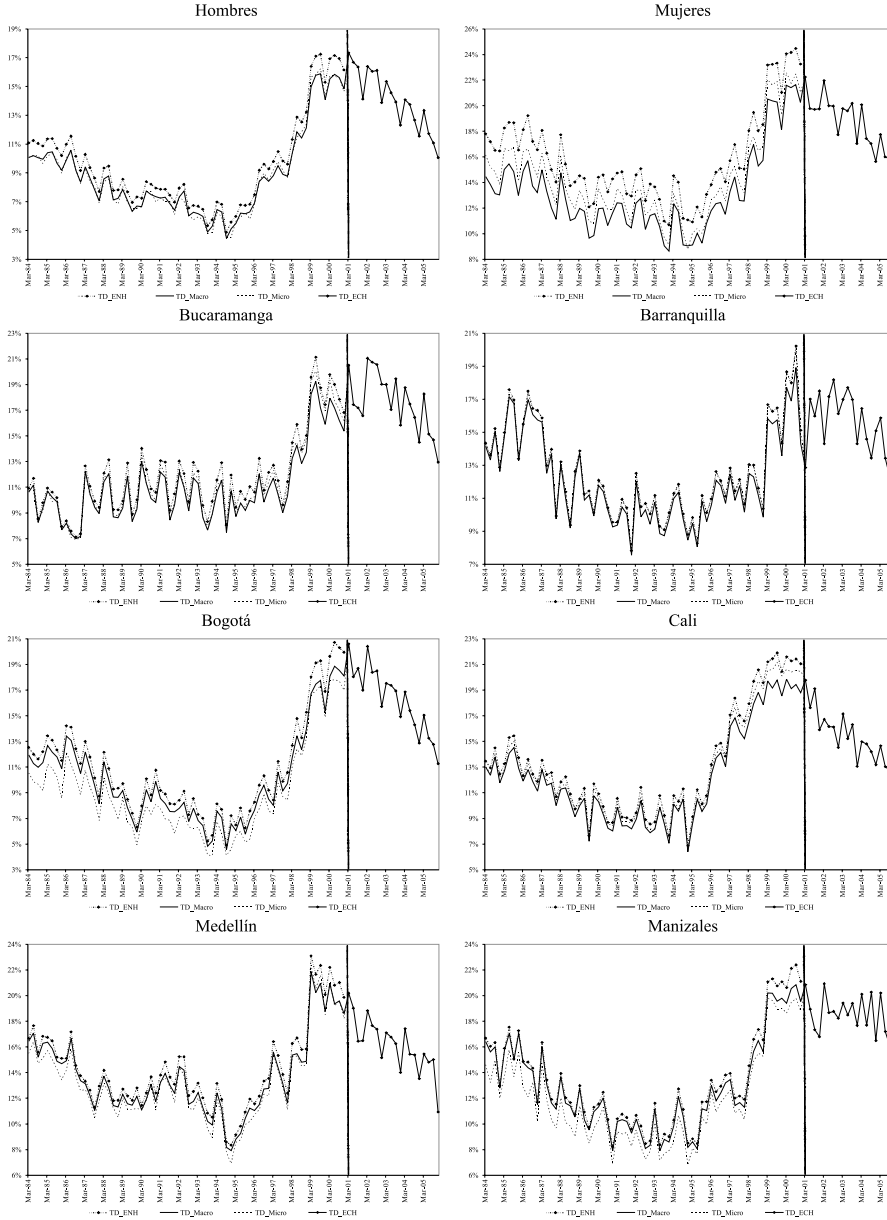
**Tabla 2.** Promedio de las principales variables del modelo *logit* para el caso de los TFSR. (...Continuación).

y	ENH		ECH	
	Total	Total	0	1
Unión libre	0,0972	0,1214	0,1217	0,0876
Casado	0,2727	0,2312	0,2317	0,1672
Viudo, separado o divorciado	0,1183	0,1405	0,1414	0,0288
Soltero	0,5118	0,5069	0,5052	0,7164
Hombre	0,3170	0,3633	0,3629	0,4130
Mujer	0,6830	0,6367	0,6371	0,5870
Edad	32,9150	36,0039	36,0889	25,3581
Escolaridad	8,2928	9,5412	9,5379	9,9490
Asistencia escolar	0,3648	0,3335	0,3319	0,5434
Sin educación	0,0444	0,0372	0,0374	0,0090
Primaria incompleta	0,1434	0,1054	0,1057	0,0745
Primaria completa	0,1838	0,1590	0,1593	0,1165
Secundaria incompleta	0,3922	0,3418	0,3408	0,4636
Secundaria completa	0,1442	0,1805	0,1807	0,1593
Educación superior	0,0920	0,1761	0,1761	0,1771
Jefe de hogar hombre	0,7693	0,6916	0,6911	0,7521
Jefe de hogar mujer	0,2307	0,3084	0,3089	0,2479
Edad jefe	49,0356	51,3274	51,3548	47,8905
Escolaridad jefe	8,1907	9,6524	9,6587	8,8561
Jefe sector agropecuario	0,0126	0,0097	0,0097	0,0127
Jefe sector industrial	0,1427	0,1062	0,1054	0,2073
Jefe sector construcción	0,0574	0,0415	0,0417	0,0248
Jefe sector comercio	0,1582	0,1487	0,1459	0,4991
Jefe sector transp. y comunic.	0,0730	0,0641	0,0642	0,0421
Jefe sector financiero	0,0484	0,0121	0,0121	0,0044
Jefe sector servicios sociales	0,1541	0,0549	0,0551	0,0407
Jefe otros sectores	0,0116	0,0998	0,1001	0,0638
Jefe TFSR	0,0519	0,0017	0,0011	0,0677
Jefe empleado	0,3701	0,2828	0,2841	0,1226
Jefe servicio doméstico	0,0060	0,0117	0,0118	0,0054
Jefe cuenta propia	0,2292	0,2224	0,2196	0,5795
Jefe patrón	0,0519	0,0461	0,0451	0,1678
Hogar sin cónyuge	0,2667	0,3315	0,3323	0,2389
Ocupados en el hogar	1,5968	1,4310	1,4253	2,1547
Desocupados en el hogar	0,3207	0,3698	0,3713	0,1766
Niños menores de 6 años	0,2029	0,1416	0,1415	0,1520
Patrón no jefe en el hogar	0,0172	0,0213	0,0209	0,0721
TGPc	0,5955	0,6488	0,6488	0,6501

Fuente: ECH y ENH. Cálculos de los autores.

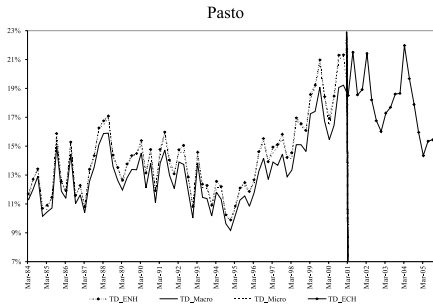
## Anexo 2

Gráfico 1. Tasa de desempleo empalmada por género y ciudad. 1984-2005.



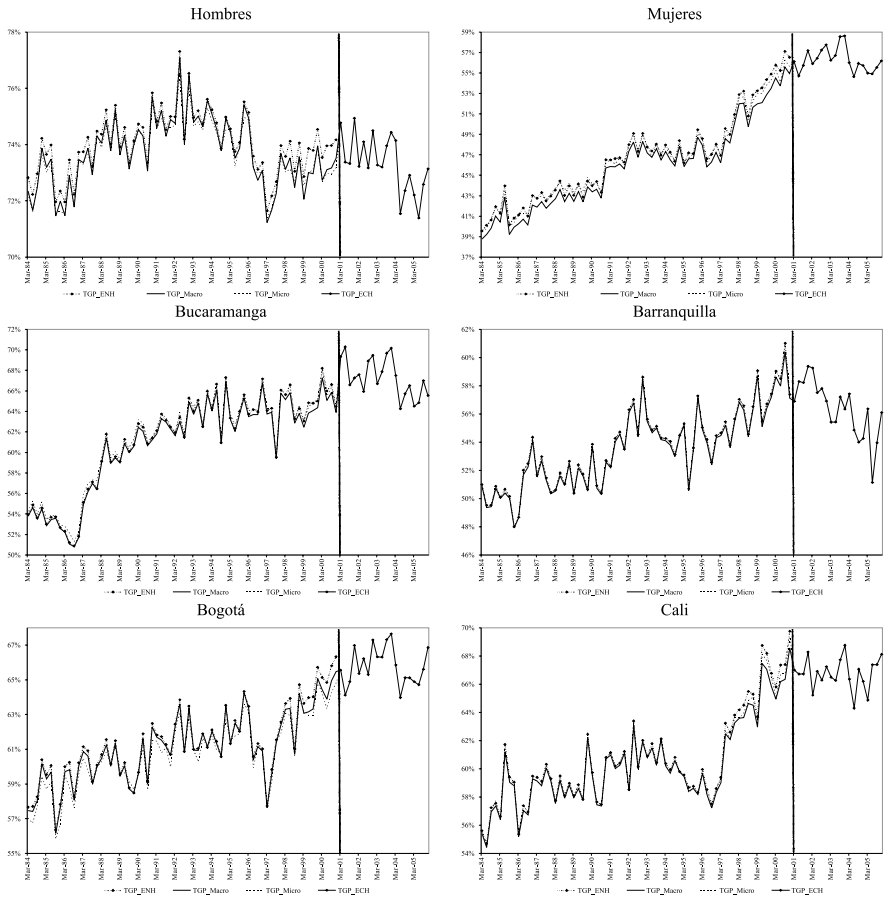
(Continúa...)

**Gráfico 1.** Tasa de desempleo empalmada por género y ciudad. 1984-2005. (...Continuación).



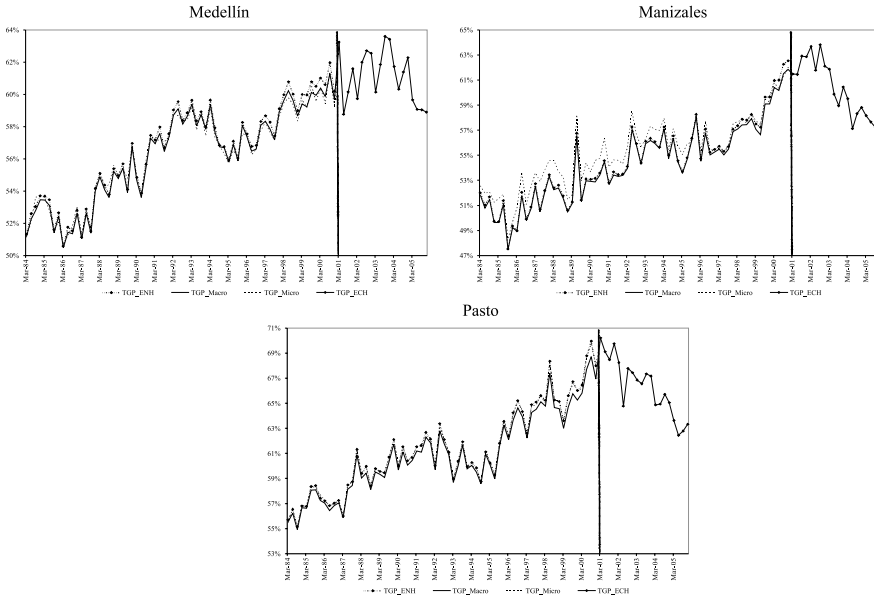
Fuente: ENH 1984-2000, ECH 2001-2005. Cálculos de los autores.

**Gráfico 2.** Tasa global de participación empalmada desagregada por género y ciudad. 1984-2005.



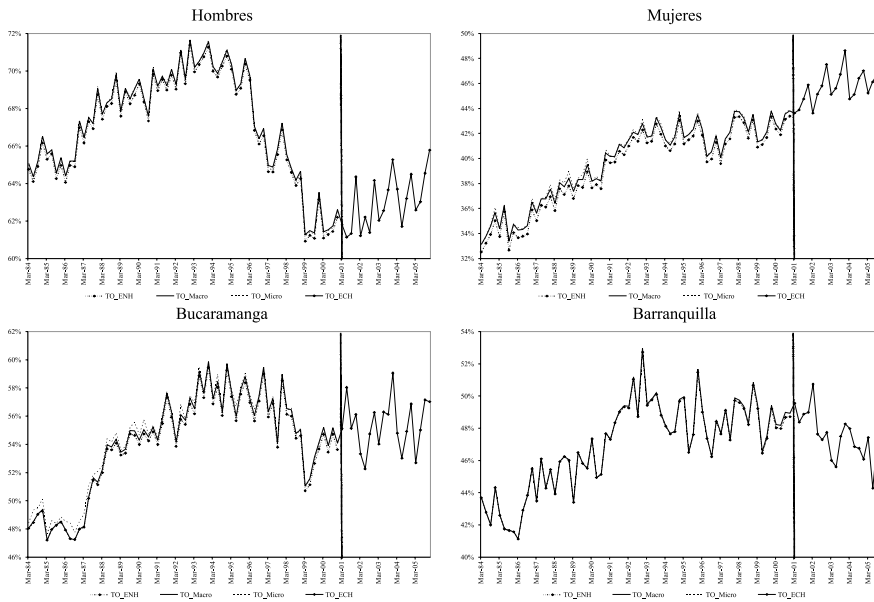
(Continúa...)

**Gráfico 2. Tasa global de participación empalmada desagregada por género y ciudad. 1984-2005. (...Continuación).**



Fuente: ECH 2001-2005. Cálculos de los autores.

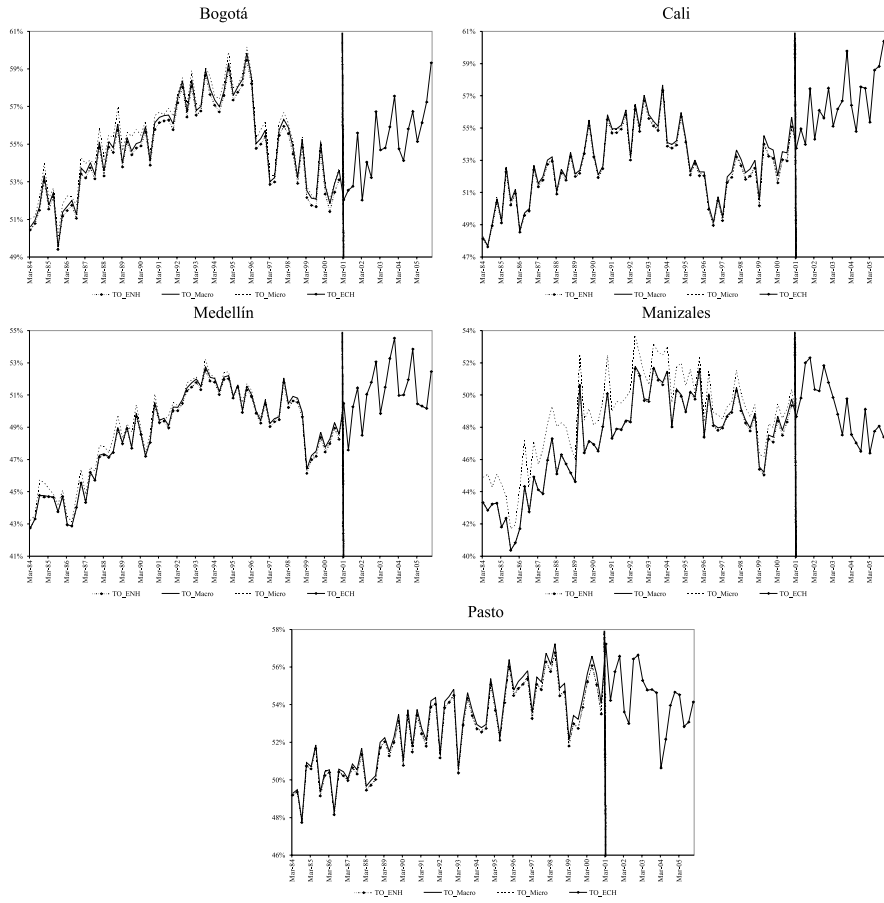
**Gráfico 3. Tasa de ocupación empalmada desagregada por género y ciudad. 1984-2005.**



(Continúa...)



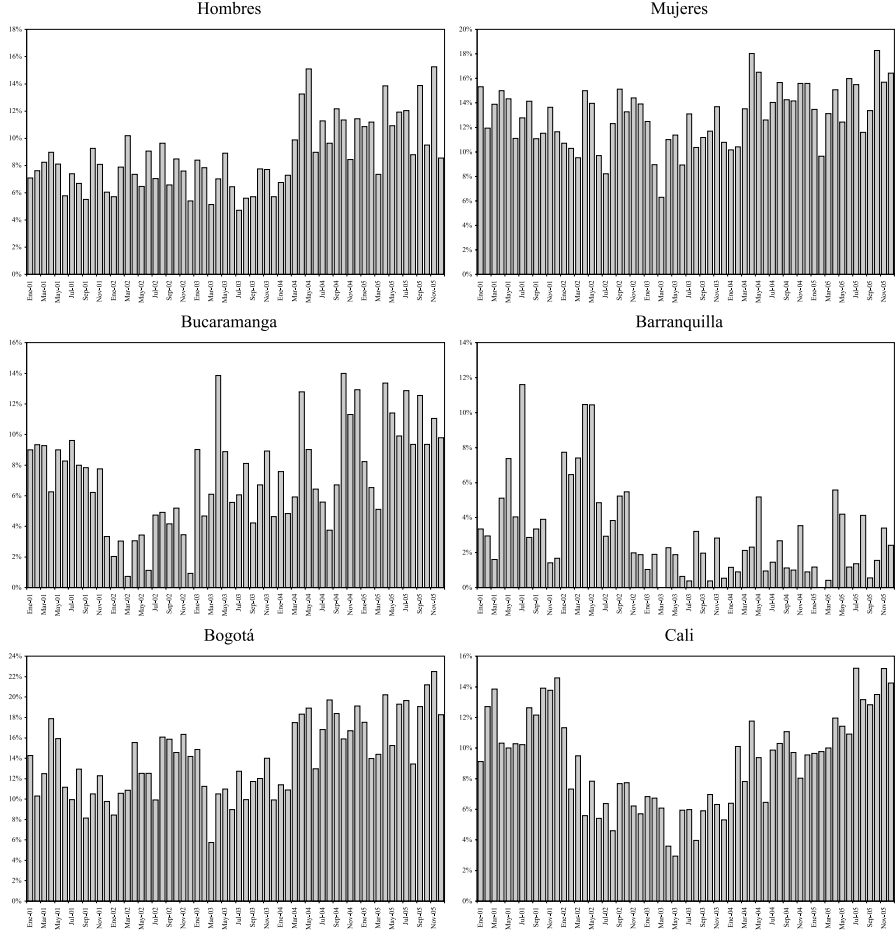
**Gráfico 3. Tasa de ocupación empalmada desagregada por género y ciudad. 1984-2005. (...Continuación).**



*Fuente:* ENH 194-2000. ECH 2001-2005. Cálculos de los autores.

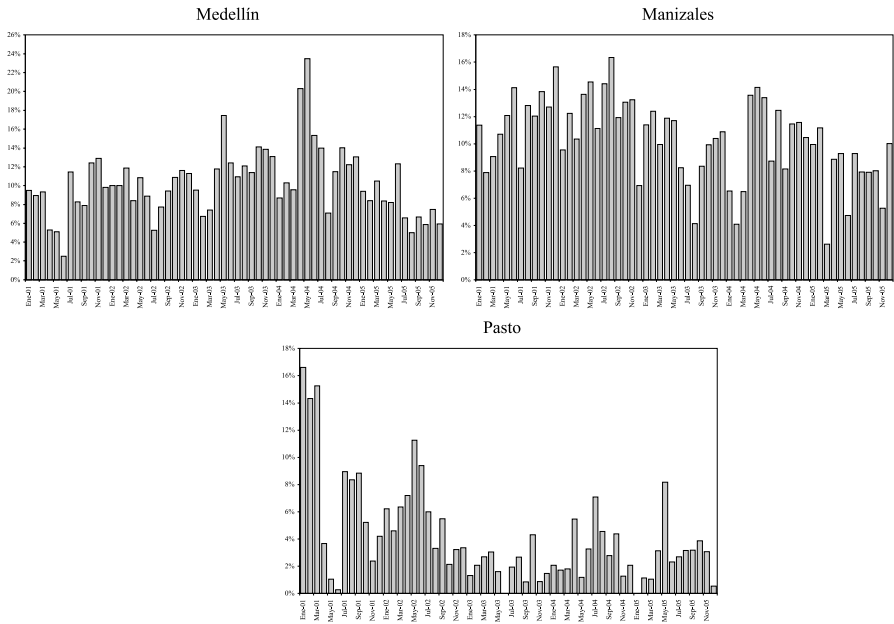
### Anexo 3

**Gráfico 4. Porcentaje de desocupados reclasificados como inactivos. 2001-2005.**



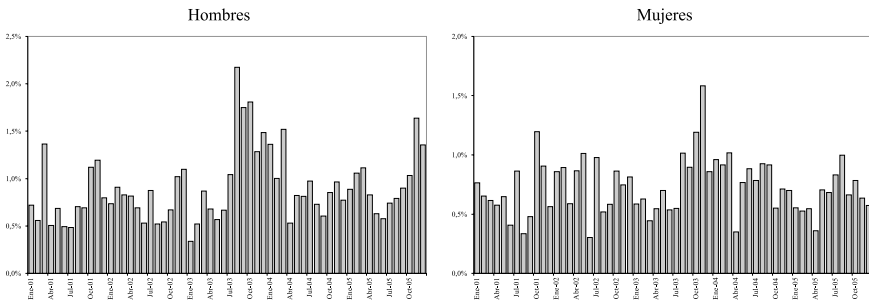
(Continúa...)

**Gráfico 4. Porcentaje de desocupados reclasificados como inactivos. 2001-2005. (...Continuación).**



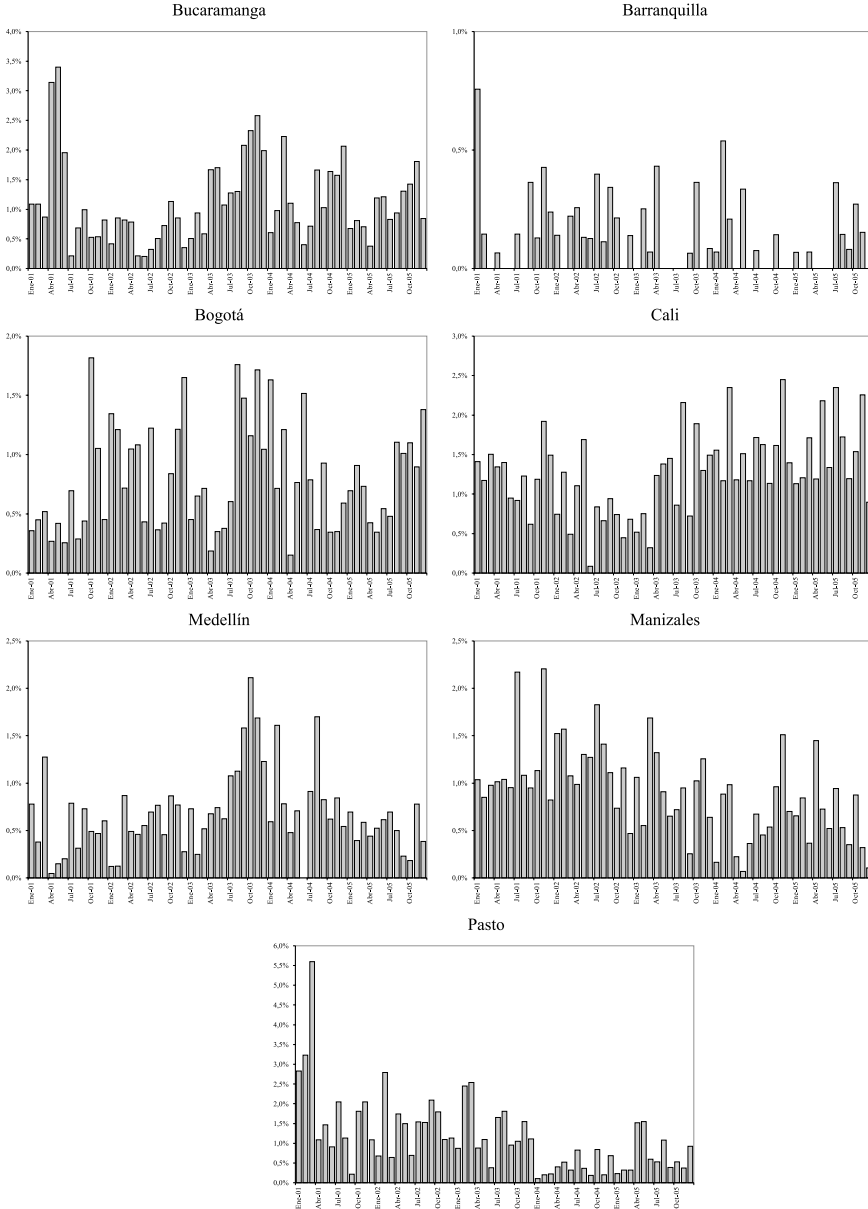
Fuente: ECH 2001-2005. Cálculos de los autores.

**Gráfico 5. Porcentaje de desocupados e inactivos reclasificados como TFSR. 2001-2005.**



(Continúa...)

**Gráfico 5. Porcentaje de desocupados e inactivos reclasificados como TFSR. 2001-2005. (...Continuación).**



Fuente: ECH 2001-2005. Cálculos de los autores.



## Anexo 4

Tabla 3. Pruebas de raíz unitaria para las series estimadas de desempleo para el total nacional y desagregadas por género y ciudad.

Desagregaciones		ADF			ADF-ERS			Estadísticos seie en diferencias			
		Estadístico	Valor-p	Rezagos	Estadístico	Valor-p	Rezagos	Rango	S.E.	Simetría	Curtois
Tasa nacional	ENH	-0,463	0,891	4	-0,753	>0,1	4	0,0623	0,0117	0,4946	3,7174
	Macro	-0,233	0,928	4	-0,596	>0,1	4	0,0595	0,0115	0,5969	3,7830
	Micro	-0,862	0,793	4	-1,053	>0,1	4	0,0649	0,0120	0,9185	4,0283
Hombres	ENH	-0,867	0,792	4	-1,024	>0,1	4	0,0512	0,0094	0,3131	3,8915
	Macro	-0,833	0,802	4	-0,998	>0,1	4	0,0469	0,0088	0,1631	3,6448
	Micro	-1,097	0,711	4	-1,200	>0,1	4	0,0520	0,0092	0,5760	3,9061
Mujeres	ENH	-0,067	0,948	4	-0,486	>0,1	4	0,0748	0,0159	0,5318	3,2613
	Macro	0,273	0,975	4	-0,266	>0,1	4	0,0738	0,0159	0,7000	3,3765
	Micro	-0,337	0,912	4	-0,686	>0,1	4	0,0795	0,0171	0,9142	3,4448
Bucaramanga	ENH	-0,602	0,862	4	-0,663	>0,1	4	0,1014	0,0216	0,0625	2,6596
	Macro	-0,722	0,833	4	-0,860	>0,1	4	0,0914	0,0195	0,1971	2,8040
	Micro	-0,778	0,818	4	-0,898	>0,1	4	0,0897	0,0194	0,4024	2,8572
Barranquilla	ENH	-1,897	0,332	1	-1,837	> 0,05 y < 0,1	1	0,1141	0,0210	0,3856	3,4397
	Macro	-1,862	0,348	1	-1,759	> 0,05 y < 0,1	1	0,1058	0,0204	0,3540	3,1831
	Micro	-1,823	0,366	1	-1,776	> 0,05 y < 0,1	1	0,1137	0,0209	0,3639	3,4636
Bogotá	ENH	-0,760	0,824	0	-0,933	>0,1	0	0,0621	0,0151	0,3435	2,1098
	Macro	-0,827	0,805	0	-0,984	>0,1	0	0,0578	0,0138	0,4087	2,2828
	Micro	-0,190	0,933	4	-0,564	>0,1	4	0,0606	0,0142	0,7475	2,7944
Cali	ENH	-0,870	0,791	5	-0,991	>0,1	5	0,0847	0,0160	-0,1219	3,1766
	Macro	-0,981	0,755	5	-1,082	>0,1	5	0,0765	0,0148	-0,0779	3,1087
	Micro	-0,966	0,745	5	-1,098	>0,1	5	0,0783	0,0146	-0,1404	3,4164
Medellín	ENH	-1,812	0,372	0	-1,741	> 0,05 y < 0,1	0	0,1058	0,0166	1,2266	7,0093
	Macro	-1,923	0,320	0	-1,745	> 0,05 y < 0,1	0	0,1001	0,0158	1,3507	7,3134
	Micro	-1,517	0,518	4	-1,423	>0,1	4	0,1087	0,0176	1,4367	7,2280
Manizales	ENH	-1,538	0,508	0	-1,478	>0,1	0	0,0801	0,0181	0,3143	2,7745
	Macro	-1,774	0,389	1	-0,986	>0,1	1	0,0785	0,0179	0,3464	2,8244
	Micro	-0,635	0,855	1	-0,841	>0,1	1	0,0756	0,0178	0,2646	2,6568
Pasto	ENH	-2,346	0,161	0	-1,991	> 0,01 y < 0,05	0	0,0812	0,0178	-0,0372	2,5895
	Macro	-2,682	0,082	0	-2,418	> 0,01 y < 0,05	0	0,0774	0,0168	0,0147	2,7214
	Micro	-2,290	0,178	0	-1,860	> 0,01 y < 0,05	0	0,0800	0,0174	-0,0831	2,5759

Fuente: cálculos de los autores.

Tabla 4. Raíz del error cuadrático medio para los pronósticos del modelo agregado y el modelo *logit*.

Período	Nacional		Hombres		Mujeres		Bucaramanga		Barranquilla	
	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>Logia</i>
2000	0,00845	0,01105	--	--	--	--	0,01403	0,01942	0,02721	0,03197
2001	0,01044	0,01056	0,00958	0,00884	0,01654	0,02285	0,01545	0,01169	0,02312	0,02554
2001-2002	0,01624	0,01570	0,01030	0,01003	0,02171	0,02732	0,02700	0,02233	0,02111	0,02297

Período	Bogotá		Cali		Medellín		Manizales		Pasto	
	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>logit</i>	Variables agregadas	<i>Logia</i>
2000	0,01060	0,00317	0,01949	0,03064	0,01499	0,01590	0,03744	0,03097	0,01871	0,03207
2001	0,01292	0,00971	0,02138	0,03275	0,02233	0,02097	0,02465	0,01780	0,01287	0,02681
2001-2002	0,01698	0,01195	0,03370	0,04607	0,02618	0,02496	0,02189	0,01547	0,02384	0,03882

Fuente: cálculos de los autores.