

XIV Encuentro de Economía Aplicada
Huelva, 2 y 3 de junio de 2011

Racionamiento de Empleo Formal e Informalidad Evidencia para Colombia

Gustavo Adolfo García*

Universitat Autònoma de Barcelona

gustavoadolfo.garcia@uab.cat

Departamento de Economía Aplicada, Campus Bellaterra,
Edificio B, 08193 Bellaterra (Cerdanyola), Telf. 935811153 - Fax 935812292

Versión preliminar
Marzo 2011

Resumen

Este trabajo pretende analizar los determinantes de la decisión de localización laboral entre la formalidad y la informalidad, teniendo en cuenta la posible existencia de una cola de empleo por los puestos de trabajo formales para el mercado laboral de Colombia. A partir del modelo de [Abowd & Farber \(1982\)](#) y especificando una regresión *switching* endógena se prueba la existencia de la cola de empleo formal y se determinan los factores a los que responde ubicarse en el sector formal/informal. Entre los principales resultados se ha encontrado que no es posible rechazar la hipótesis de cola de empleo formal, lo cual es evidencia del racionamiento de empleos formales que existe en el mercado laboral colombiano. Este racionamiento divide el proceso de colocación laboral en dos etapas: en la primera etapa el trabajador debe decidir si juntarse o no a la cola por un puesto formal y en la segunda los empleadores deben decidir qué trabajador seleccionar de la cola para que ocupe la vacante formal. Los resultados muestran que la probabilidad ubicarse en la cola por un empleo formal aumenta con la prima salarial que ofrece el sector formal y el nivel de responsabilidad que tiene el individuo en el hogar (jefe de hogar y estar casado). Respecto a la decisión de los empleadores, se ha encontrado que ésta depende positivamente del nivel de cualificación, el grado de responsabilidad, el compromiso y el buen historial laboral del individuo, y negativamente de una fase de desempleo previa.

Palabras clave: Informalidad laboral, racionamiento de empleo formal, regresión *switching* endógena, Colombia

Códigos JEL: C35, J23, J24, J42

*Agradezco los comentarios de los profesores José Luis Roig y Josep Luís Raymond (UAB) a versiones preliminares de este trabajo. Se agradece la financiación de la Generalitat de Cataluña, becas predoctorales FI.

1. Introducción

Los países en desarrollo se caracterizan por una importante heterogeneidad en los mercados laborales. Esta heterogeneidad se expresa en la coexistencia de un sector moderno o formal, altamente productivo, con buenas remuneraciones, altos requerimientos de capital y mano de obra calificada; con otro sector tradicional o informal, menos desarrollado, intensivo en mano de obra, con bajos niveles educativos, de baja productividad y al margen de las regulaciones laborales.

Desde una perspectiva estructuralista, se plantea que esta dualidad es producto de las limitaciones que presenta el sector moderno de la economía para generar los suficientes puestos de trabajo. En este sentido, el remanente de mano de obra que no logra ubicarse en el sector moderno o formal se ve obligado a emplearse en puestos de trabajo de menor calidad, y, en peor situación, otra proporción cae en el desempleo (Uribe & Ortiz 2006, Ortiz et al. 2010). Como se observa en el cuadro 1, este doble mal de escasez de empleos con baja calidad de los existentes, no es exclusiva de una región. Se observa que para una serie de regiones y países en desarrollo la informalidad genera entre el 50 y 60 % de los empleos y entre el 30 y 40 % del producto total, mientras que los desempleados representan alrededor de un 8 % de la población activa, con casos extremos como África del Norte y Colombia en donde las tasas de desempleo superan el 10 %.

Teniendo en cuenta este mayor peso de las actividades informales en la economía y siendo la calidad del empleo un elemento importante de la calidad de vida y el bienestar socioeconómico, además de un mecanismo para contrarrestar las desigualdades sociales, este trabajo pretende analizar las características y determinantes del proceso de colocación laboral entre la formalidad y la informalidad, a partir de una metodología alternativa que asume la existencia de un racionamiento de puestos de trabajo formales. La literatura a nivel nacional¹ se ha basado en modelos que no tienen en cuenta el racionamiento de empleos formales, con lo cual ubicarse en el sector formal o informal es una decisión exclusivamente del individuo que depende únicamente de sus características o preferencias. La metodología que se propone seguir se basa en un enfoque de cola de empleo y tiene en cuenta en el proceso de localización laboral el criterio que tienen los empleadores en la selección para ocupar los limitados puestos de trabajo. El proceso de colocación, entonces, se divide en dos mecanismos: uno es la elección que hacen los trabajadores por juntarse a la cola de empleo por un puesto de trabajo formal y el otro es el mecanismo que aplican los empleadores para escoger a los trabajadores de la cola de empleo para ocupar la vacante.

El análisis de los determinantes de la localización laboral se basa en la estimación

¹ Véase por ejemplo Nuñez (2002), Ribero (2003), Uribe et al. (2006), Roldan & Ospina (2009), Bernal (2009) y Mondragón-Vélez et al. (2010). En estos trabajos se hacen estimaciones univariantes (logit o probit) para analizar la probabilidad de ser informal.

Cuadro 1. Niveles de empleo y producto informal por regiones

Región	Informalidad	Desempleo	PIB informal
Asia	70 %	5.6 %	35 %
África del Norte	47 %	10.3 %	36 %
América Latina	54 %	7.3 %	41 %
Argentina	53 %	7.9 %	25 %
Brasil	51 %	7.9 %	40 %
Chile	36 %	7.8 %	20 %
México	50 %	4.9 %	30 %
Colombia	58 %	13 %	39 %

Fuente: Informalidad: [Jütting & De Laiglesia \(2009\)](#) y corresponde a diferentes períodos: Asia (meridional y sudoriental) 1995-99; África del Norte 2000-07; América Latina, Argentina y Chile 1995-99; Brasil y México 2000-07; y para Colombia se tomó de la GEIH-DANE segundo trimestre de 2009. Desempleo: [OIT \(2009a\)](#) y [OIT \(2009b\)](#) y la información es para el año 2008. PIB informal: [Schneider & Klinglmair \(2004\)](#), corresponde al año 2000. La definición de informalidad que se sigue corresponde a trabajadores que no están cubiertos por la seguridad social. Schneider y Klinglmair tienen en cuenta tanto actividades ilegales como legales que evaden impuestos para contabilizar el PIB informal.

de un modelo de regresión *switching* endógeno. Este modelo además de tener en cuenta el doble mecanismo que se presenta en el proceso de localización laboral, incluye como determinantes adicionales las predicciones no sesgadas de las primas salariales por ubicarse en la formalidad. Las estimaciones se realizan para el mercado laboral de Colombia, el cual como se observa en el cuadro 1, presenta un importante desequilibrio tanto cuantitativo como cualitativo: considerables niveles de informalidad junto con una alta tasa de desempleo. La información con la que se lleva a cabo el análisis proviene de los microdatos de la Gran Encuesta Integrada de Hogares (GEIH) para todos los meses de 2009. El análisis se realiza a nivel urbano para las trece principales áreas metropolitanas de Colombia.

Entre los resultados más importantes se ha encontrado que existe un racionamiento de empleos formales y tal racionamiento ha afectado el 16 % de los trabajadores que se ubican en la informalidad. Esta restricción de empleos formales es evidencia de la segmentación laboral que sufre el mercado laboral colombiano, algo que se viene presentando desde la década de los noventa ([Uribe et al. 2007](#)). En cuanto a los determinantes de la localización en la formalidad/informalidad se ha encontrado que la primera de las decisiones, la de buscar un empleo formal, se encuentra principalmente influenciada por las características personales, de capital humano, del hogar y las primas salariales que implican trabajar en la formalidad. Por su parte, los factores que más influyen las ofertas de empleo son la dotación de capital humano, el compromiso, la responsabilidad en el empleo y el historial laboral de los individuos.

Este trabajo se organiza de la siguiente forma: además de esta introducción, en

la sección 2 se expone el modelo econométrico que se sigue, luego en la sección 3 se describe la estrategia de estimación y se proponen algunos test para validar los resultados. En la sección 4 se detalla la base de datos utilizada y se muestran algunas estadísticas descriptivas con referencia a la población formal e informal. En la sección 5 se muestran los resultados obtenidos para finalizar en la sección 6 con las conclusiones.

2. El modelo econométrico

Desde el punto de vista neoclásico las decisiones en el mercado laboral se encuentran determinadas por las características del trabajador, es decir, las condiciones de la oferta. En este sentido, los agentes escogen voluntariamente sus opciones: participación o inactividad, empleo o desempleo, trabajo formal o informal, etc., siguiendo un proceso de maximización del bienestar sujeto a una serie de restricciones propias del individuo, como son el tiempo disponible, la riqueza inicial, la dotación de capital humano, entre otras. Desde este enfoque la estructura de la economía y las características de la demanda laboral no son tenidas en cuenta en los factores que determinan la ocupación o el desempleo.

En una visión alternativa del mercado laboral puede plantearse que los trabajadores sí escogen, sí deciden, como plantean los neoclásicos, pero la amplitud de las opciones se encuentra determinada por las restricciones que se imponen del lado de la demanda. En esta visión no se supone que la demanda laboral es amplia y diversa, como si se hace en la neoclásica, por el contrario existen restricciones en los empleos y sobre todo en los formales, ya que el sector moderno sólo demanda el trabajo formal que necesita (Archibald 1977, Dickens & Lang 1985, Lang & Dickens 1987, Uribe & Ortiz 2006).

Al existir restricciones en los puestos de trabajo las decisiones de querer trabajar no implican directamente que el individuo sea contratado, es decir que la probabilidad de desear un empleo en un determinado sector no es igual a la probabilidad de tener un empleo en tal sector. Teniendo en cuenta además la posibilidad de un exceso de demanda por un tipo de empleo (una “cola” de empleo), observar a un trabajador empleado en un determinado puesto de trabajo depende de las reglas de racionamiento de los empleos que utilizan los empleadores así como las preferencias de los trabajadores por dicho puesto.

Para modelar esta situación de racionamiento de puestos de trabajo formales, se va seguir el modelo de cola de empleo de Abowd & Farber (1982) y la extensión de Mengistae (1999). En el primer trabajo se desarrolla un modelo para analizar los determinantes del estatus sindical de los trabajadores en el que se tiene en cuenta

la posibilidad de una cola de empleo para los puestos de trabajos sindicalizados, mientras que en el segundo utilizando el modelo de Abwod y Farber, se analiza la existencia de una cola de empleo en el sector público y las diferencias salariales que existen entre este último sector y el privado. Con este mismo enfoque en el análisis de las restricciones en el mercado laboral y la existencia de colas de empleo para puestos de trabajo formales se encuentra los trabajos de [Dickens & Lang \(1985\)](#), [Maloney \(1998\)](#), [Soares \(2004\)](#), [Co et al. \(2005\)](#), [Puentes & Contreras \(2009\)](#), y [Contreras et al. \(2008\)](#).

Se tiene, entonces, que desde este punto de vista, en el mercado laboral los trabajadores eligen libremente el sector en el que quieren trabajar, sin embargo tal elección se encuentra restringida por las opciones de empleo que existan en el mercado. Las restricciones estructurales en la economía limitan el número de puestos de trabajo formal y al ser estos preferidos por la mayoría de los trabajadores ya que ofrecen mejores condiciones laborales (estabilidad, seguridad social, posibilidades de promoción y recompensa por la acumulación de capital humano) se constituye una “cola” de solicitantes de estos puestos de trabajo ([Dickens & Lang 1985](#), [Lang & Dickens 1987](#)). La existencia de una cola de empleo para el sector formal relaja el supuesto de que la probabilidad de querer estar empleado en el sector formal sea igual a la probabilidad de conseguir un empleo formal. Bajo este enfoque observar a un individuo en la formalidad depende de dos procesos de decisión: la decisión del trabajador de hacer parte de la cola para acceder a un empleo formal y la decisión del empleador de contratar al trabajador que se encuentra en la cola.

Analicemos la primera de las decisiones: hacer parte de la cola de empleo para los puestos formales. La decisión de los trabajadores de hacer parte de la cola por los puestos de trabajo formales depende de la utilidad que les represente dicho empleo. Un determinante importante en esta decisión y en la maximización de la utilidad es la prima salarial que le ofrece estar en un sector o en otro. Sea entonces: W_{1i} el salario hora que gana un individuo i en el sector formal y W_{2i} el salario en el sector informal. También definamos a U_{1i} la utilidad máxima que puede lograr el individuo i en el sector formal y U_{2i} la utilidad máxima que alcanza en el sector informal. Se tiene, entonces, que el individuo i preferirá un empleo formal y por tanto hará cola por un puesto de trabajo de este tipo si $V_{1i}^* = U_{1i} - U_{2i} > 0$. Asumiendo que V_{1i}^* depende linealmente de la prima salarial que ofrece el sector formal ($\ln W_{1i} - \ln W_{2i}$) y de un conjunto de características del individuo y de su entorno laboral y familiar (X_{1i}), se puede escribir la siguiente relación que representa la ecuación de la cola de empleo:

$$V_{1i}^* = \alpha_{1i}(\ln W_{1i} - \ln W_{2i}) + X_{1i}'\beta_{1i} + u_{1i} \quad (1)$$

donde α_{1i} es una constante, β_{1i} es un vector de parámetros incluido un término constante y u_{1i} es un término de error aleatorio distribuido normal con media cero y varianza $\sigma_{u_1}^2$.

Ya que V_{1i}^* no es observable directamente y sólo es posible observar su signo, se

puede definir una variable binaria I_{1i} que resuma la disposición de un individuo para obtener un empleo formal, esto es:

$$\begin{aligned} I_{1i} &= 1 \quad \text{si} \quad V_{1i}^* = \alpha_{1i}(\ln W_{1i} - \ln W_{2i}) + X'_{1i}\beta_{1i} + u_{1i} > 0 \\ I_{1i} &= 0 \quad \text{si} \quad V_{1i}^* = \alpha_{1i}(\ln W_{1i} - \ln W_{2i}) + X'_{1i}\beta_{1i} + u_{1i} \leq 0 \end{aligned} \quad (2)$$

Entonces, I_{1i} es un variable indicador igual a 1 si el trabajador se encuentra en la cola e igual a cero si el trabajador prefiere no hacer la cola.

La segunda decisión, la de los empleadores que seleccionan a los trabajadores que se encuentran en la cola para ocupar la vacante formal, también puede modelarse a partir de una variable latente V_{2i}^* . Lang y Dickens (1988, pp. 76) plantean que la existencia de una cola de empleo supone que los empleadores del sector formal hacen un ranking de trabajadores de acuerdo a sus cualidades y seleccionan al mejor trabajador disponible de la cola. Sea entonces Q_n las cualidades del individuo n y Q_N las cualidades del resto de individuos en la cola. Se tiene así, que un empleador escogerá a un individuo de la cola para llenar la vacante, siempre y cuando las cualidades de este individuo superen la de todos los otros individuos en la cola, es decir que $V_{2i}^* = Q_n - Q_N > 0$. Se supone que dichas cualidades dependen de un conjunto de características observables del individuo (X_{2i}) y del grado de productividad que tendría el individuo. Una forma que tienen los empleadores de contabilizar el nivel de productividad (además del capital humano de los individuos, el cual ya se encuentra incluido en X_{2i}), es determinando el salario esperado que obtendría el individuo en el sector formal dada sus características. Así, individuos con un mayor salario esperado pueden indicar altos niveles de productividad y, por tanto, un mayor producto, que es lo que busca el maximizar el empleador. Con los anteriores supuestos la ecuación de ser seleccionado de la cola de empleo para un puesto formal, tiene la siguiente forma:

$$V_{2i}^* = \alpha_{2i}E(\ln W_{1i}|Z_{1i}, V_{1i}^* > 0) + X'_{2i}\beta_{2i} + u_{2i} \quad (3)$$

donde i representa al grupo de trabajadores que están en la cola; α_{2i} es una constante; $E(\ln W_{1i}|Z_{1i}, V_{1i}^* > 0)$ es el salario esperado que un trabajador podría obtener en el sector formal dadas sus características personales (Z_{1i}) y dado que se encuentra en la cola; β_{2i} es un vector de parámetros incluyendo un término constante; y u_{2i} es un término de error aleatorio distribuido normal con media cero y varianza $\sigma_{u_2}^2$.

Al igual que en la ecuación de cola de empleo, se define una variable binaria I_{2i} que recoge el signo de V_{2i}^* y resume el proceso de selección que hace el empleador al escoger un trabajador de la cola, esto es:

$$\begin{aligned} I_{2i} &= 1 \quad \text{si} \quad \alpha_{2i}E(\ln W_{1i}|Z_{1i}, V_{1i}^*) + X'_{2i}\beta_{2i} + u_{2i} > 0 \\ I_{2i} &= 0 \quad \text{si} \quad \alpha_{2i}E(\ln W_{1i}|Z_{1i}, V_{1i}^*) + X'_{2i}\beta_{2i} + u_{2i} \leq 0 \end{aligned} \quad (4)$$

Entonces, I_{2i} es una variable indicador igual a 1 si el trabajador es escogido de la

cola para llenar la vacante disponible en el sector formal e igual a 0 si no es seleccionado.

Junto con las ecuaciones (1) y (3) el sistema se completa con las ecuaciones salariales del sector formal e informal:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + v_{1i} \quad \text{si } V_{1i}^* > 0 \text{ y } V_{2i}^* > 0 \quad (5)$$

$$\ln W_{2i} = Z'_{2i}\gamma_2 + v_{2i} \quad \text{en otro caso} \quad (6)$$

donde, Z_{ji} es un vector de características individuales y laborales, γ_1 y γ_2 son vectores de parámetros, v_{1i} y v_{2i} son normales *iid* con media cero y varianzas $\sigma_{v_1}^2$ y $\sigma_{v_2}^2$, respectivamente. Mengistae (1999) muestra que, asumiendo una distribución conjunta de los términos de error $(u_{1i}, u_{2i}, v_{1i}, v_{2i})$, la matriz de covarianza tiene la forma

$$\begin{bmatrix} \sigma_{u_1}^2 & \sigma_{1u_2} & \sigma_{1v_1} & \sigma_{1v_2} \\ & \sigma_{u_2}^2 & \sigma_{2v_1} & \sigma_{1v_2} \\ & & \sigma_{v_1}^2 & \sigma_{v_1v_2} \\ & & & \sigma_{v_2}^2 \end{bmatrix},$$

donde, $Cov(u_{1i}, v_1) = \sigma_{1v_1}$; $Cov(u_{1i}, v_2) = \sigma_{1v_2}$; $Cov(u_{2i}, v_1) = \sigma_{2v_1}$; $Cov(u_{2i}, v_2) = \sigma_{2v_2}$; $Cov(v_{1i}, v_{2i}) = \sigma_{v_1v_2}$; y $Cov(u_{1i}, u_{2i}) = \sigma_{1u_2}^2$.

La estimación por MCO de las anteriores ecuaciones salariales sin tener en cuenta las decisiones de localización del sector (formal o informal) que hacen los trabajadores y las decisiones de selección de los empleadores, puede generar sesgos. $\ln W_{1i}$ y $\ln W_{2i}$ son variables censuradas ya que la primera se observa sólo para los trabajadores formales y la segunda sólo para los informales, por lo que los términos de perturbación v_{1i} y v_{2i} tienen una distribución truncada y las estimaciones por MCO no serían consistentes por la existencia de un problema de sesgo de selección (Heckman 1979, Huguet 1996).

Para identificar completamente el sistema de ecuaciones falta deducir el salario esperado que un trabajador podría obtener en el sector formal estando en la cola (es decir, la expresión $E(\ln W_{1i} | Z_{1i}, V_{1i}^* > 0)$ de la ecuación (3)). Utilizando las ecuaciones (1) y (5) se tiene que

$$E(\ln W_{1i} | Z_{1i}, V_{1i}^* > 0) = Z'_{1i}\gamma_1 + \sigma_{1v_1}\lambda_{1i} \quad (7)$$

donde λ_{1i} es análoga a la razón inversa de Mills y su estructura depende de si ρ es igual o diferente de cero. En la siguiente sección se desarrolla su expresión.

El modelo completo, entonces, se encuentra conformado por las ecuaciones (1), (3), (5) y (6). Este sistema es endógeno ya que en las ecuaciones de localización de

² En la sección 3.1 este parámetro se referencia como ρ . Mengistae (1999) parte suponiendo que este parámetro es cero ya que su estimación presenta una estructura secuencial.

empleo (ecuaciones (1) y (3)) se incluyen los salarios (ecuaciones (5) y (6)) como determinantes adicionales, a su vez las ecuaciones de salarios se encuentran determinados por las decisiones de localización laboral de los trabajadores y las decisiones de selección de los empleadores. A este tipo de modelos se les llama “endogenous switching regression model” (Maddala & Nelson 1975). En la siguiente sección se discute la estrategia de estimación de este tipo de modelos, el cual va a depender del grado de observabilidad de los individuos que están en la cola y el supuesto sobre el proceso de decisiones: si es secuencial o simultáneo.

3. Estrategia de estimación y los test de existencia de cola de empleo

Para la estimación del modelo *switching* endógeno se va seguir un procedimiento en tres etapas. En una primera etapa se estima la forma reducida del probit bivariante (donde no se incluye el diferencial de salarios entre sectores) y se toman los residuales para aproximar las esperanzas de v_{1i} y v_{2i} condicionadas a u_{1i} y u_{2i} . Estas aproximaciones, que no son otra cosa que las razones inversas de Mills, podrían en una segunda etapa entrar a la ecuación salarial de cada uno de los sectores y corregir el sesgo de selección. Finalmente, en la tercera etapa se incluye como regresor adicional en la ecuación de “juntarse a la cola de empleo” (JC) el diferencial de salarios estimados en la segunda etapa. De igual forma se incluye en la ecuación de “ser escogido de la cola de empleo” (EDC) el salario que un trabajador podría obtener en el sector formal, estimado de las ecuaciones salariales.

Para la identificación del sistema es necesario plantear algunas restricciones de exclusión en el modelo. Por un lado, para poder identificar los parámetros de las ecuaciones de “juntarse a la cola” y “ser escogido de la cola”, se debe incorporar variables que afecten la primera decisión pero que no influya sobre la segunda, o viceversa. Por otro lado, para la identificación de los parámetros de las ecuaciones salariales, es necesario incluir, al menos, una variable que determine los salarios pero que no afecte las decisiones de “juntarse a la cola” o “ser escogido de la cola”.

Otro aspecto importante a tener en cuenta en la estimación del probit bivariante es la observabilidad parcial de los datos³. Las encuestas de hogares por lo regular sólo ofrecen información sobre la situación actual de los trabajadores, es decir si es informal o formal, pero no informa sobre el proceso de decisiones que lleva a un individuo a ubicarse en un tipo de empleo. Esto implica que no es posible observar los diferentes tipos de trabajadores informales que surgen de las decisiones: los trabajadores que

³ Véase Abowd & Farber (1982), Meng & Schmidt (1985) y Tunali (1986) para una explicación más detallada sobre la observabilidad parcial

deciden ser informales y no hacen cola por un empleo formal, y los trabajadores que aún estando en la cola no son elegidos y se ubican en la informalidad como opción de empleo. En términos de las ecuaciones descritas en la anterior sección, no se dispone de información de las elecciones de “juntarse a la cola” (I_{1i}) ni “ser escogido de la cola” (I_{2i}), solo es posible observar el producto de estas dos decisiones ($I_1 = I_{1i}I_{2i}$). Si este producto es igual a 1, el trabajador está en el sector formal y si es igual a 0 se encuentra en el sector informal.

Para superar esta falta de información Poirier (1980) y Abowd & Farber (1982) plantean utilizar la misma variable dependiente para las ecuaciones de “juntarse a la cola” y “ser escogido de la cola” y hacer una diferenciación entre las características que afectan cada una de las ecuaciones. De aquí la importancia de las restricciones de exclusión para poder identificar los efectos.

Un aspecto que diferencia la metodología de Poirier de la de Abowd y Farber es el supuesto sobre el proceso de elección. El primer autor asume que ambas decisiones (juntarse a la cola y ser escogido de la cola) son tomadas simultáneamente, mientras que los segundos suponen que el proceso de decisión es secuencial. Según Maddala (1983) la elección de uno u otro modelo depende del propósito que se busque. Si el propósito del análisis es examinar los factores que influyen en la decisión de los empleadores de contratar a un tipo específico de trabajador, el modelo simultáneo sería más adecuado, mientras que si el interés es analizar todo el proceso de colocación de los trabajadores entre la formalidad y la informalidad, teniendo en cuenta tanto los factores que afectan las preferencias de los trabajadores y de los empleadores, el modelo secuencial será más apropiado. Este último enfoque para nuestro caso parece ser más adecuado, sin embargo, en la medida de lo posible se van a seguir las dos metodologías para observar la consistencia de los resultados.

3.1. Estimación

Como se mencionó, la estimación del modelo depende del supuesto que se haga sobre el proceso de decisiones: si es secuencial o simultáneo. En el modelo de Abowd y Farber se asume que las decisiones se toman de manera secuencial, lo cual supone que dichas decisiones son independientes. Lo anterior implica que la función de distribución conjunta de los términos de perturbación de las ecuaciones de “juntarse a la cola” y “ser escogido de la cola” (u_{1i} y u_{2i} , respectivamente) siguen una distribución normal bivalente con media cero, varianza unitaria y covarianza igual cero. Bajo esta estructura la función de verosimilitud con observabilidad parcial a ser maximizada es (Meng & Schmidt 1985, Tunali 1986):

$$L_1 = \prod_{I=1} [\Phi(X'_{1i}\beta_1)\Phi(X'_{2i}\beta_2)] \prod_{I=0} [1 - \Phi(X'_{1i}\beta_1)\Phi(X'_{2i}\beta_2)] \quad (8)$$

donde, Φ es la función de distribución acumulada normal, X_{1i} son las variables explicativas de la ecuación de juntarse a la cola y X_{2i} son las variables que explican la ecuación de ser escogido de la cola. Como en la primera etapa se estima la forma reducida del probit bivalente, también se incluyen como variables explicativas las variables que se asumen sólo afectan los salarios.

Por su parte, el modelo de Poirier supone que el proceso de localización entre la formalidad e informalidad se hace de manera simultánea, lo que implica que la correlación entre los términos de perturbación de las dos ecuaciones es diferente de cero. Bajo este último supuesto se tiene que la función de verosimilitud a maximizar tiene la siguiente estructura (Meng & Schmidt 1985, Tunali 1986):

$$L_2 = \prod_{I=1} [\Phi_2(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho)] \prod_{I=0} [1 - \Phi_2(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho)] \quad (9)$$

donde Φ_2 es la función de distribución acumulada bivalente normal y ρ es la covarianza entre u_{1i} y u_{2i} , suponiendo que $\sigma_{u_1}^2 = \sigma_{u_2}^2 = 1$.

Nótese que si se supone que no existe correlación entre los términos de perturbación de las dos ecuaciones, ρ será igual a cero en la ecuación (9) y el modelo simultáneo se convierte en el modelo secuencial.

Siguiendo con la segunda etapa de la estimación, planteada al principio de esta sección, se debe estimar las ecuaciones salariales corrigiendo el sesgo que implica el censuramiento de los datos generado por el proceso de localización entre la formalidad e informalidad. Para corregir el anterior sesgo de selección se sigue el procedimiento propuesto por Maddala (1983, pp. 278-283), el cual es una extensión del método de estimación en dos etapas de Heckman-Lee (Heckman 1976,9, Lee 1978). Asumiendo una distribución normal en las perturbaciones de la ecuación salarial de los trabajadores formales, el primer momento de la distribución truncada viene dado por la siguiente expresión (Tunali 1986, Mohanty 2001):

$$E(v_{1i}|I_{1i} = 1, I_{2i} = 1) = \sigma_{1v_1}\lambda_{1i} + \sigma_{2v_1}\lambda_{2i} \quad (10)$$

donde λ_{1i} y λ_{2i} son análogas a las razones inversas de Mills en el contexto univariante. Los valores de estas razones dependen de la existencia o no de correlación entre las decisiones de “juntarse a la cola” y “ser escogido de la cola”, esto es si o no $Cov(u_{1i}, u_{2i}) = \rho = 0$. Si se supone que la anterior expresión es diferente de cero se tiene que (Tunali 1986):

$$\lambda_{1i} = \frac{\phi(X'_{1i}\beta_1)\Phi(X'_{2i}\beta_2 - \rho X'_{1i}\beta_1) / \sqrt{1 - \rho^2}}{\Phi_2(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho)} \quad (11)$$

$$\lambda_{2i} = \frac{\phi(X'_{2i}\beta_2)\Phi(X'_{1i}\beta_1 - \rho X'_{2i}\beta_2) / \sqrt{1 - \rho^2}}{\Phi_2(X'_{1i}\beta_1, X'_{2i}\beta_2, \rho)} \quad (12)$$

Ahora, si se supone independencia entre las decisiones entonces $\rho = 0$, por lo que las anteriores expresiones se reducen a:

$$\lambda_{1i} = \frac{\phi(X'_{1i}\beta_1)}{\Phi(X'_{1i}\beta_1)} \quad (13)$$

$$\lambda_{2i} = \frac{\phi(X'_{2i}\beta_2)}{\Phi(X'_{2i}\beta_2)} \quad (14)$$

Así, la estimación consistente de la ecuación salarial para los trabajadores ubicados en el sector formal, se obtiene incorporando como regresores adicionales a λ_{1i} y λ_{2i} , es decir estimando la siguiente ecuación:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + \sigma_{1v_1}\lambda_{1i} + \sigma_{2v_1}\lambda_{2i} + \epsilon_{1i} \quad (15)$$

donde $\epsilon_{1i} = v_1 - \sigma_{1v_1}\lambda_{1i} - \sigma_{2v_1}\lambda_{2i}$ y $E(\epsilon_{1i}|I_i = 1) = 0$. Siguiendo el método Heckman-Lee una aproximación a las expresiones de λ_{1i} y λ_{2i} son las estimaciones que provienen de los probit bivariantes (simultáneo o secuencial), lo cual arroja estimaciones consistentes de γ_1 . Esto es, aplicando MCO a la siguiente expresión:

$$\ln W_{1i} = Z'_{1i}\gamma_1 + \sigma_{1v_1}\hat{\lambda}_{1i} + \sigma_{2v_1}\hat{\lambda}_{2i} + \eta_{1i} \quad (16)$$

donde $\eta_{1i} = \epsilon_{1i} + \sigma_{1v_1}(\lambda_{1i} - \hat{\lambda}_{1i}) + \sigma_{2v_1}(\lambda_{2i} - \hat{\lambda}_{2i})$.

Al seguir el método en dos etapas de Heckman-Lee para la ecuación salarial de los trabajadores informales ($I_i = 0$) surgen problemas de identificación de los tipos de informales y por tanto las reglas de selección muestral son más complicadas que para el caso de los trabajadores formales. Para el modelo de Poirier (caso simultáneo), en principio se pueden diferenciar dos tipos de trabajadores informales. Unos que están en la cola de empleo pero no son seleccionados por los empleadores ($E(v_{2i}|I_{1i} = 1 \text{ e } I_{2i} = 0)$) y los otros que han decidido no juntarse a la cola ($E(v_{2i}|I_{1i} = 0)$). Además, existe otro tipo de trabajador informal no claramente definido (Tunali (1986, pp. 274); Soares (2004, pp. 11, 17), el cual aún no juntándose a la cola podrían ser seleccionado de ella para un puesto formal ($E(v_{2i}|I_{1i} = 0 \text{ e } I_{2i} = 1)$)⁴. Con este problema de identificación de los tipos de informales no es posible calcular las razones inversas de Mills y, por tanto, hacer la corrección de selección a la ecuación salarial de los informales con el modelo de Poirier.

⁴ Las reglas de selección muestral y por tanto la construcción de la razones inversa de Mills son posibles en el caso de los trabajadores formales, es decir cuando $I_i = I_{1i}I_{2i} = 1$. Éste producto sólo es igual a 1 cuando el trabajador ha hecho la cola y es elegido de ella para ocupar la vacante formal ($I_{1i} = 1 \text{ e } I_{2i} = 1$). Sin embargo, para el caso de los trabajadores informales, es decir cuando $I_i = I_{1i}I_{2i} = 0$, existen tres combinaciones en las cuales el producto es cero. Estas son: no hacer la cola y no ser elegido de la cola ($I_{1i} = 0 \text{ e } I_{2i} = 0$), haber hecho la cola pero no ser elegido de ella ($I_{1i} = 1 \text{ e } I_{2i} = 0$) y no hacer la cola y ser elegido de la cola ($I_{1i} = 0 \text{ e } I_{2i} = 1$). Esta última regla de selección no es clara por lo que las razones inversas de Mills no pueden ser definidas bajo el modelo de Poirier.

Para el caso del modelo secuencial (modelo de Abowd y Farber), Mengistae (1999, pp. 11-12) ha propuesto una solución para hacer la corrección del sesgo de selección a la ecuación salarial de los informales. Dicho autor muestra que el salario esperado para los trabajadores informales ($I_i = 0$) puede ser estimado como el promedio ponderado entre el salario de los trabajadores que son informales por no juntarse a la cola de empleo ($I_{1i} = 0$) y el salario esperado de los trabajadores que resultan ser informales al no ser seleccionado de la cola de empleo habiendo hecho dicha cola ($I_{1i} = 1$ e $I_{2i} = 0$). Esto es (Mengistae 1999):

$$(1 - \pi)E(W_{2i}^1 | I_{1i} = 0) + \pi E(W_{2i}^2 | I_{1i} = 1 \text{ e } I_{2i} = 0), \quad (17)$$

con π como la proporción de trabajadores informales que están en la cola de empleo y no son escogidos y $(1 - \pi)$ como la proporción de trabajadores informales que no están en la cola. La anterior expresión puede escribirse como:

$$\ln W_{2i} = Z'_{2i} \gamma_2 + \sigma_{1v_2} \lambda_{3i} + \delta_1 \lambda_{1i}^* + \delta_2 \lambda_{4i} + \epsilon_{2i} \quad (18)$$

donde, $\lambda_{3i} = -\phi(X'_{1i} \beta_1) / [1 - \Phi(X'_{1i} \beta_1)]$; $\lambda_{4i} = -\phi(X'_{2i} \beta_2) / [1 - \Phi(X'_{2i} \beta_2)]$; $\lambda_{1i}^* = \lambda_{1i} - \lambda_{3i}$; $\delta_1 = \pi \sigma_{1v_2}$; $\delta_2 = \pi \sigma_{2v_2}$; $\epsilon_{2i} = (1 - \pi) \epsilon_{2i}^1 + \pi \epsilon_{2i}^2$; ϵ_{2i}^1 es el término de error de la ecuación salarial para $I_{1i} = 0$; ϵ_{2i}^2 es el término de error de la ecuación salarial para $I_{1i} = 1$ e $I_{2i} = 0$.

Estimaciones consistentes de γ_2 , σ_{1v_2} , δ_1 y δ_2 son obtenidas aplicando MCO a la ecuación:

$$\ln W_{2i} = Z'_{2i} \gamma_2 + \sigma_{1v_2} \hat{\lambda}_{3i} + \delta_1 \hat{\lambda}_{1i}^* + \delta_2 \hat{\lambda}_{4i} + \eta_{2i} \quad (19)$$

donde $\eta_{2i} = \epsilon_{2i} + \sigma_{1v_2} (\lambda_{3i} - \hat{\lambda}_{3i}) + \delta_1 (\lambda_{1i}^* - \hat{\lambda}_{1i}^*) + \delta_2 (\lambda_{4i} - \hat{\lambda}_{4i})$ y las razones inversa de Mills son aproximadas a partir del modelo probit bivalente Abowd-Farber.

Con el procedimiento en dos etapas de Heckman-Lee se logran estimaciones consistentes de los parámetros, sin embargo si el sesgo de selección existe los errores estándar de dichas estimaciones no serán consistentes. Tunali (1986) y Mengistae (1999) proponen una metodología para corregir los errores estándar. En este trabajo se sigue la metodología del último autor para hacer la corrección.

En la tercera etapa del modelo *switching* endógeno se estiman las ecuaciones de “juntarse a la cola” y “ser escogido de la cola” con un probit bivalente estructural. Como ya se mencionó, en la primera ecuación se incluye como regresor adicional la diferencia estimada de los salarios y en la segunda ecuación se adiciona los salarios estimados del sector formal. El probit bivalente estructural sólo se podrá estimar para el modelo Abowd-Farber ya que para el caso del modelo Poirier no es posible obtener las estimaciones sin selectividad de los salarios del sector informal.

Un último aspecto a tener en cuenta en la estimación del modelo *switching* endógeno es el que se refiere a las restricciones de exclusión. Existen dos conjuntos de

restricciones. Las primeras tienen que ver con las restricciones que se deben imponer para identificar separadamente las ecuaciones de “juntarse a la cola” y “ser escogido de la cola”. Y las segundas son las que hacen posible la identificación de las ecuaciones salariales.

Para las primeras restricciones se supone que los factores que inciden sobre la decisión de juntarse a la cola por un empleo formal se encuentran estrechamente relacionados con las características personales y sobre todo con las familiares o del hogar. Por su parte, las decisiones de contratación de los empleadores se ven más afectadas por la dotación de capital humano, como la educación y la experiencia, el estatus ocupacional anterior, el compromiso por el trabajo y la afiliación sectorial del empleo. Se tiene, entonces, que las variables excluidas de la segunda ecuación e incluidas en la primera, son: la antigüedad en el actual trabajo y las variables relacionadas con el hogar, como el número de niños, el estatus ocupacional de otros miembros del hogar y el nivel educativo del hogar.

Veamos los supuestos que hay detrás de la inclusión o exclusión de algunas de las variables antes mencionadas (para más detalle de estos supuestos véase Soares (2004)). De la ecuación de “ser escogido de la cola” se han excluido las variables experiencia y antigüedad. La primera se excluye para evitar problemas de multicolinealidad, ya que al ser experiencia potencial esta construida a partir de la educación. Respecto a la segunda, se supone que los empleadores se fijan en la experiencia que tendría a lo largo de toda la vida laboral y no en un puesto en particular (Abowd & Farber 1982, Farber 1983, Soares 2004). En cuanto a las variables de posición dentro del hogar (si es jefe de hogar o no) y el estado civil (soltero y casado), su inclusión intenta medir dos efectos diferentes. En la ecuación de “ser escogido de la cola” estas variables representan el compromiso por el trabajo, en la medida en que las responsabilidades que asumen las cabezas del hogar pueden estar relacionadas con un mayor compromiso en el empleo; y en la ecuación de “juntarse a la cola” representan las tradicionales características del individuo, que según sea su posición y obligaciones en la estructura familiar así mismo se vera su compromiso por ubicarse en un determinado empleo.

Por su parte, con la inclusión de las variables del hogar en la primera ecuación se quiere contabilizar los costos que implican para el individuo ubicarse en la formalidad o en la informalidad. El número de niños es una variable proxy de la carga laboral en el hogar, sobre todo para las mujeres y los jefes de hogar (hombre o mujer). Para las primeras sería más beneficioso un empleo informal que implique flexibilidad horaria para poder distribuir su tiempo con mayor facilidad entre el cuidado de los niños y el trabajo. Por su parte, para los jefes de hogar, la presencia de niños implica un alto costo el estar desempleado, por lo que es más probable ubicarse en la informalidad.

En cuanto a las segundas restricciones de exclusión para identificar las ecuaciones salariales se asume que las características del hogar y el estatus laboral anterior no tienen ningún efecto sobre los niveles salariales de los formales e informales. También

se supone para la identificación que la afiliación sectorial del empleo solamente afecta los salarios, con lo cual estas variables se excluyen del probit bivalente estructural.

3.2. Existencia de la cola de empleo

Para probar la existencia de cola por los empleos formales se va a seguir los test propuestos por [Abowd & Farber \(1982\)](#) y [Mengistae \(1999\)](#). Estos tests además de validar la robustez del modelo bivalente respecto al univalente, muestra evidencia de las barreras que existen para acceder a un puesto de trabajo formal. En otras palabras, la existencia de la cola de empleo implica que no todos los trabajadores que quieren trabajar en la formalidad lo pueden hacer debido a las restricciones que imponen sus propias características y las limitaciones en los puestos de trabajo formales que existan.

[Abowd & Farber \(1982\)](#) proponen evaluar la existencia de una cola de empleo a partir de las estimaciones del modelo probit bivalente de la primera etapa. En ausencia de una cola de empleo los parámetros de la ecuación de “ser escogido de la cola” se restringen a cero con excepción del término constante. A partir de las estimaciones restringidas y no restringidas del modelo probit bivalente se prueba la hipótesis nula de no existencia de cola a partir de un estadístico de razón de verosimilitud (LR). Otra prueba que se podría hacer, opuesta al escenario de no cola, es el caso de la existencia de una cola universal. Una cola universal es equivalente a la restricción de que todos los parámetros de la ecuación de “juntarse a la cola” son cero con excepción del término constante. De igual forma que en el primer test, con un estadístico de razón de verosimilitud se contrasta la hipótesis nula de existencia de una cola universal.

[Mengistae \(1999\)](#) propone tests basados en las estimaciones condicionadas de las ecuaciones salariales. Los test parten de contrastar hipótesis alrededor del parámetro π de la ecuación (18), el cual representa la proporción de trabajadores informales que han sido racionados del sector formal. Probar la no existencia de una cola de empleo implica contrastar la hipótesis nula de $\pi = 0$, mientras que el contraste de $\pi = 1$ implica corroborar la existencia de una cola universal. La primera prueba se realiza con un test de significancia conjunta de δ_1 y δ_2 ($H_0: \delta_1 = \delta_2 = 0$), y para la segunda se prueba la restricción de $\sigma_{1v_2} = \delta_1$ a partir de un estadístico F. La limitación de este test es que no se puede hacer para el modelo de Poirier, ya que, como se mencionó en la sección anterior, no existe una estimación para la ecuación salarial de los informales.

Los test de Abowd y Farber, y Mengistae siguen diferentes metodologías y pueden no llevar a las mismas conclusiones. En la sección 5.4 se muestran los resultados de ambos tests.

4. Base de datos y análisis descriptivo

Los datos usados en este trabajo provienen de la Gran Encuesta Integrada de Hogares (GEIH) de Colombia para todos los meses del año 2009. Esta encuesta es realizada por el Departamento Administrativo Nacional de Estadística (DANE) y proporciona información básica acerca del tamaño y estructura de la fuerza laboral (empleo, desempleo e inactividad) para las 13 principales áreas metropolitanas de Colombia⁵.

Se han eliminado las observaciones que presentan datos *missing* en alguna de las variables de tal forma que no haya problemas en la convergencia de los probit bivariantes. El número de observaciones resultante es de 154613 trabajadores, que representan en la muestra ampliada 8.6 millones de registros.

Para definir a los informales se ha utilizado una medida que se enmarca dentro de la teoría institucionalista (De Soto 1987, Portes 1995). Bajo esta teoría la informalidad se asocia con la evasión de las normas legales, en el sentido que las cargas fiscales e ineficiencias gubernamentales incentivan el incumplimiento de las reglas institucionales y, por tanto, la informalidad. Se tiene, entonces, que los informales son aquellos trabajadores que no tienen acceso al sistema de seguridad social en salud y pensión. Esta medida es la más comúnmente utilizada dentro de este enfoque⁶ y a diferencias de las definiciones estructuralistas, que se basan en el tamaño de la empresa y medidas de productividad, presentan menor heterogeneidad⁷. Siguiendo la anterior definición, se tiene que hay una muestra de 98713 trabajadores informales, lo cuales representa en los datos expandidos cerca de 5 millones de trabajadores, que se traduce en una tasa de informalidad urbana de 58 %.

Ahora pasamos a caracterizar los formales e informales teniendo en cuenta aspectos personales, del hogar y del empleo. De los cuadros 2 y 3 se observa que los trabajadores formales son en promedio más educados que los informales y la diferencia es de 3.5 años. De los primeros la mayoría tiene estudios secundarios (42 %) y superiores (47 %), mientras que los segundos se concentran en estudios primarios (30 %) y secundarios (52 %). Respecto al sexo se observa que los informales presentan una mayor equiparación entre hombres y mujeres que los formales. Para la característica jefatura del hogar se nota que los informales son en su mayoría no jefes de hogar (56 %), lo cual puede indicar que la búsqueda de ingresos secundarios del hogar se hace en la informalidad.

⁵ A saber Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Manizales, Pereira, Cúcuta, Pasto, Villavicencio, Cartagena, Montería e Ibagué.

⁶ Véase por ejemplo Saavedra & Chong (1999), Maloney (2004), Bosch & Esteban-Preteel (2009), Lehmann & Pignatti (2007), Chong et al. (2007), Bosch & Maloney (2008), Galli & Kucera (2008), Perry et al. (2007) y Jütting & De Laiglesia (2009).

⁷ Para una discusión más amplia de las diferentes definiciones de informalidad puede verse Khamis (2009).

En cuanto a la experiencia laboral (cuadro 3) se nota que los informales presentan más años medios: la diferencia entre informales y formales es de 5.2 años medios de experiencia laboral. Del lado de la antigüedad en el actual empleo no se observan marcadas diferencias, tanto formales como informales tienen un promedio de 6 años de antigüedad. En los ingresos laborales se tiene que los formales doblan en remuneraciones a los informales, al parecer el mayor número de años de experiencia de los informales no es un indicio de una mayor productividad de este tipo de trabajadores.

Cuadro 2. Características personales y del empleo

	Formales	Informales	Formales/Informales	Total ocupados
Educación				
Sin estudios	0.31 %	2.46 %	12.60 %	1.55 %
Estudios primarios	9.76 %	29.14 %	33.49 %	20.94 %
Estudios secundarios	42.48 %	51.98 %	81.72 %	47.96 %
Estudios superiores	47.45 %	16.41 %	289.15 %	29.54 %
<i>Total</i>	100 %	100 %	100 %	100 %
Sexo				
Mujer	43.81 %	46.99 %	93.23 %	45.65 %
Hombre	56.19 %	53.01 %	106.00 %	54.35 %
<i>Total</i>	100 %	100 %	100 %	100 %
Jefatura del hogar				
No Jefe	53.20 %	55.77 %	93.39 %	54.68 %
Jefe	46.80 %	44.23 %	105.81 %	45.32 %
<i>Total</i>	100 %	100 %	100 %	100 %
Estado civil				
Soltero	54.08 %	44.26 %	115.97 %	44.19 %
Casado	55.92 %	55.74 %	100.32 %	55.81 %
<i>Total</i>	100 %	100 %	100 %	100 %
Estatus ocupacional anterior al actual empleo				
Empleado	82.60 %	71.60 %	115.36 %	76.25 %
Desempleado	17.40 %	28.40 %	61.27 %	23.75 %
<i>Total</i>	100 %	100 %	100 %	100 %
Sector económico				
Industria, electricidad, gas, agua	21.33 %	16.49 %	129.35 %	18.53 %
Construcción	3.64 %	7.84 %	46.43 %	6.06 %
Comercio y hostelería	19.50 %	38.47 %	50.69 %	30.44 %
Transporte y comunicación	9.29 %	10.10 %	91.98 %	9.75 %
Seguros, finanzas, admon pública, educación	32.91 %	11.24 %	292.79 %	20.41 %
Servicios personales, domésticos u otros	13.34 %	15.87 %	84.06 %	14.80 %
<i>Total</i>	100 %	100 %	100 %	100 %

Fuente: Cálculos propios a partir de la GEIH 2009.

Nota: Las estadísticas están expandidas con los factores de ponderación anual.

Del lado de las variables del hogar (cuadro 3 y 4) se tiene que los trabajadores informales presentan una mayor probabilidad de vivir en hogares con menores niveles educativos. En cuanto a la presencia de niños en el hogar, se observa que una alta proporción de informales viven en hogares con niños en edades escolares (entre 6 y 17 años). Un factor también importante en la informalidad es la presencia de más informales en el hogar. Como se observa, cerca del 59 % de los informales viven en hogares en donde hay más informales. De igual forma se tiene que la presencia de desempleados en el hogar afecta las decisiones de formalizarse: el 17 % de los informales viven en hogares con desempleados.

La distribución por rama de actividad (cuadro 2) muestra que los trabajadores formales están principalmente concentrados en el sector financiero y público (33 %), y en el sector de la industria (21 %), mientras que los trabajadores informales presentan prevalencia en el sector del comercio y hostelería (39 %) y servicios personales (16 %). Por último, el estatus ocupacional anterior muestra que cerca de una tercera parte de los informales venían de una fase de desempleo antes del actual empleo.

Cuadro 3. Ingreso laboral y nivel educativo

	Formales		Informales		Total ocupados	
	Media	D.E	Media	D.E	Media	D.E
Edad	36.4	10.7	38.2	13.6	37.4	12.5
Años de educación	12.0	4.0	8.5	4.1	10.0	4.4
Años de experiencia (potencial) laboral	19.4	11.9	24.6	15.3	22.4	14.2
Año de antigüedad en el actual empleo	6.1	7.0	6.2	11.6	6.1	9.9
Años de educación en el hogar (otros miembros)	5.5	2.7	4.7	2.5	5.1	2.6
Ingreso laboral mensual	1,164,440	1,748,133	592,354	1,929,509	842,019	1,874,140
Ingreso laboral mensual real por hora	5524.4	8127.6	3197.2	11962.1	4212.8	10526.4

Fuente: Cálculos propios a partir de la GEIH 2009.

D.E: Desviación Estándar

Experiencia = Edad-años de educación-5

Nota: Los ingresos laborales mensuales reales por hora se encuentran medidos en pesos colombianos de 2008. Se utilizó el IPC del último mes de cada trimestre y para cada ciudad, en el caso de Ibagué se tomó el IPC de Pereira, por ser ciudades con características comunes. Los ingresos laborales mensuales se encuentran en pesos colombianos de 2009. En los años de educación en el hogar se descuenta los propios años de educación del individuo, ya que se encuentran contabilizados en la variable años de educación. La tasa de cambio peso-euro para diciembre de 2009 fue de 2935 pesos colombianos por euro (fuente Banco de la República). Las estadísticas están expandidas con los factores de ponderación anual.

La anterior caracterización muestra que los informales tienen una media de 38 años, con un nivel educativo que no supera la secundaria, la mayoría son hombres, casados y no jefes de hogar. Los hogares en los que viven los informales tienen una alta presencia de menores en edades escolares, con un clima educativo precario (4.7 años medios de educación en el hogar) y en donde hay otros miembros del hogar en el desempleo o en igual situación de informalidad. Respecto al empleo se ha encontrado que los informales ganan un 49 % menos de salarios mensual que los formales y más de la mitad se ubican en el sector de comercio, hostelería y servicios personales. La informalidad, entonces, es un sector con precarias condiciones laborales, en las que abunda mano de obra poco calificada y se ubica en sectores de bajo desarrollo técnico.

Cuadro 4. Características del hogar

	Formales	informales	Total ocupados
Presencia de niños de 0-2 años	9.66 %	9.70 %	9.68 %
Presencia de niños de 3-5 años	11.10 %	12.21 %	11.74 %
Presencia de niños de 6-10 años	19.61 %	22.04 %	21.01 %
Presencia de niños de 11-17 años	27.14 %	33.91 %	27.14 %
Presencia de formales en el hogar (otros miembros)	47.67 %	30.93 %	38.01 %
Presencia de informales en el hogar (otros miembros)	39.06 %	58.67 %	50.37 %
Presencia de desempleados en el hogar (otros miembros)	16.36 %	17.75 %	17.16 %

Fuente: Cálculos propios a partir de la GEIH 2009.

Nota: Las estadísticas están expandidas con los factores de ponderación anual.

5. Resultados

5.1. Probit univariante: ausencia de una cola de empleo

La estimación del probit con ausencia de una cola de empleo ofrece una buena descripción de las variables a incluir en los determinantes de la elección que hacen los trabajadores por el sector laboral. Los resultados de esta estimación se muestran en el cuadro 5. Se observa que un mayor nivel educativo incrementa la probabilidad de ubicarse en la formalidad y los estudios superiores son los que mayor efecto tienen. Dicha probabilidad también aumenta con los años de experiencia y antigüedad pero a una tasa decreciente para la primera variable. El coeficiente de la variable sexo fue positivo y significativo, lo que muestra cierto grado de discriminación en el proceso de localización entre la formalidad e informalidad. Del lado de la jefatura del hogar y el estado civil las estimaciones muestran que ser jefe de hogar o estar casado hace más probable que un trabajador se localice en la formalidad. Estos efectos positivos, como ya se comentó, puede reflejar que las mayores responsabilidades del hogar se asocian a una mayor necesidad por un buen puesto de trabajo o buena señal en cuanto a compromiso laboral para ser contratado en un puesto formal.

Las variables del hogar muestran que la presencia de niños, sobre todo con edades preescolares (3-5 años) y mayores (11-17 años), genera un efecto negativo en la probabilidad de localizarse en la formalidad. Mientras que las variables asociadas con un mayor nivel educativo, número de formales y desempleados en el hogar generan un efecto positivo sobre la formalidad.

En cuanto a la variable de estatus ocupacional anterior, se tiene que una etapa de desempleo previa al actual empleo genera un efecto negativo en la consecución de un empleo formal, en comparación a estar ocupado (categoría base). Finalmente, la asociación sectorial del empleo muestra que ubicarse en cualquier otro sector diferente

del comercial, hostelería y construcción (categoría base), aumenta la probabilidad de localizarse en un empleo formal.

Cuadro 5. Estimación probit univariante

(Y = 1 formal, 0 informal)

	Coef.	t	Efectos marg.	z
Constante	-1.779***	-41.83		
<i>Características personales</i>				
Estudios primarios	0.247***	6.12	0.092***	5.99
Estudios secundarios	0.676***	16.81	0.241***	17.34
Estudios superiores	1.303***	31.68	0.481***	35.06
Experiencia	0.023***	21.11	0.008***	21.19
Experiencia ²	-0.0007***	-31.28	-0.0002***	-31.54
Antigüedad	0.007***	7.50	0.002***	7.50
Sexo (hombre=1)	0.122***	14.88	0.044***	14.91
Jefe de hogar (jefe de hogar=1)	0.269***	30.33	0.098***	30.34
Casado (casado=1)	0.078***	9.47	0.028***	9.49
<i>Características del hogar</i>				
# niños 0-2 años	-0.082***	-6.96	-0.030***	-6.97
# niños 3-5 años	-0.102***	-9.56	-0.037***	-9.56
# niños 6-10 años	-0.079***	-10.77	-0.029***	-10.77
# niños 11-17 años	-0.124***	-23.2	-0.045***	-23.22
# de formales en el hogar	0.249***	40.57	0.090***	40.54
# desempleados en el hogar	0.026***	3.50	0.009***	3.50
Educación promedio del hogar	0.014***	7.71	0.005***	7.71
<i>Características del empleo</i>				
Desempleado antes del actual empleo	-0.348***	-35.83	-0.120***	-37.74
Industria, electricidad., gas y agua	0.501***	49.33	0.190***	48.21
Transporte y comunicación	0.174***	13.86	0.065***	13.55
Seguros, finanzas, admon pública, educación	0.784***	73.73	0.299***	73.95
Servicios personales, domésticos u otros	0.297***	26.58	0.112***	25.87
N	154613			
Log L	-81479.80			
Pseudo R2	0.19			
AIC	163003.60			
BIC	163222.47			

Fuente: Cálculos propios a partir de la GEIH 2009.

*** p<0.01, ** p<0.05, * p<0.1

Errores estándar robustos a la heteroscedasticidad. Categorías base: Educación: Sin estudios; Estado civil: soltero; Sector: comercio, hostelería y construcción.

5.2. Probit bivariante forma reducida: existencia de una cola de empleo

La anterior forma univariante supone que el proceso de localización entre la formalidad e informalidad omite el mecanismo de selección que hacen los empleadores y,

por tanto, el posible racionamiento de puestos formales que exista. Como se ha mencionado una especificación más adecuada para describir el proceso de localización laboral en presencia de restricciones en el empleo, es la estimación de un modelo bivariante. Las estimaciones de los modelos probit bivariante con observabilidad parcial en su forma secuencial (modelo Abowd-Farber) y simultánea (modelo Poirier) se muestran en el cuadro 6.

Cuadro 6. Estimaciones probit bivariante forma reducida
($Y_j = 1$ formal, 0 informal; $j = JC, EDC$)

	Modelo Abowd-Farber (Secuencial)				Modelo Poirier (Simultáneo)			
	JC		EDC		JC		EDC	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z
Constante	-1.641***	-24.65	1.385***	5.62	-1.665***	-22.89	2.125***	9.42
<i>Características personales</i>								
Estudios primarios	0.131***	2.08	0.359***	2.97	0.170**	2.47	0.092	0.74
Estudios secundarios	0.548***	8.63	0.606***	5.12	0.615***	8.86	0.119	0.92
Estudios superiores	1.324***	20.58	0.695***	5.69	1.400***	19.86	0.121	0.86
Experiencia	0.025***	19.56			0.023***	19.11		
Experiencia ²	-0.0007***	-31.75			-0.0007***	-28.92		
Antigüedad	0.023***	23.58			0.019***	19.85		
Sexo (hombre=1)	0.758***	20.28	-2.035***	-9.70	0.950***	21.73	-2.103***	-11.28
Jefe de hogar (jefe de hogar=1)	0.231***	16.96	0.111***	4.97	0.203***	15.17	0.100***	5.18
Casado (casado=1)	0.008	0.67	0.195***	8.79	-0.013	-1.11	0.189***	9.89
<i>Características del hogar</i>								
# niños 0-2 años	-0.102***	-6.80			-0.094***	-6.75		
# niños 3-5 años	-0.130***	-9.66			-0.117***	-9.31		
# niños 6-10 años	-0.101***	-10.93			-0.090***	-10.41		
# niños 11-17 años	-0.153***	-22.5			-0.140***	-21.46		
# de formales en el hogar	0.302***	38.34			0.291***	38.77		
# desempleados en el hogar	0.037***	3.88			0.030***	3.40		
Educación promedio del hogar	0.015***	6.80			0.015***	7.30		
<i>Características del empleo</i>								
Desempleado antes del actual empleo	-0.381***	-29.7	-0.305***	-14.06	-0.352***	-25.72	0.228***	-11.09
Industria, electricidad., gas y agua	0.335***	19.52	0.676***	21.29	0.313***	16.62	0.496***	16.93
Transporte y comunicación	0.234***	9.20	0.046	1.48	0.232***	9.08	-0.006	-0.19
Seguros, finanzas, admon pública, educación	0.711***	42.09	0.692***	27.2	0.689***	37.15	0.537***	19.02
Servicios personales, domésticos u otros	0.194***	11.98	0.602***	15.77	0.176***	9.96	0.466***	13.86
ρ					-0.576***	-14.78		
N		154613				154613		
Log L		-80682.71				-80636.63		
AIC		161433.41				161343.27		
BIC		161771.67				161691.47		

Fuente: Cálculos propios a partir de la GEIH 2009

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Categorías base: Educación: Sin estudios; Sector: comercio, hostelería y construcción

JC: ecuación de "juntarse a la cola"; EDC: ecuación de "ser escogido de la cola"

Se observan interesantes patrones para las ecuaciones de "juntarse a la cola" (JC) y "ser escogido de la cola" (EDC). Se tiene que un mayor nivel educativo incrementa la probabilidad de juntarse a la cola para un puesto de trabajo formal, teniendo un mayor efecto los niveles altos de educación. Respecto a la ecuación de ser escogido de la cola, se nota igualmente que la educación ejerce un efecto positivo, pero este patrón sólo se presenta en el modelo secuencial.

Las otras variables de capital humano, la experiencia y la antigüedad, presentan un efecto positivo y decreciente tanto en la probabilidad de juntarse a la cola como en la de ser escogido de ella para un empleo formal.

Respecto a las características personales los resultados muestran diferentes efectos. Para la variable de género se tiene que ser hombre aumenta la probabilidad de juntarse a la cola de empleo, mientras que las mujeres presentan mayor probabilidad de ser seleccionadas de la cola. Estos efectos pueden deberse a las diferentes preferencias y mecanismos que utilizan hombres y mujeres en la búsqueda de empleo. En la característica de ser jefe de hogar, se ha obtenido efectos positivos tanto en la decisión del trabajador por buscar un empleo formal como en el proceso de selección de los empleadores. Como ya se ha mencionado, la característica de jefe de hogar se encuentra relacionada con mayor compromiso en el empleo, con lo cual los empleadores lo ven positivamente en el proceso de contratación. Por último, se observa que entre un individuo soltero (categoría base) y uno casado no existen diferencias en el proceso de búsqueda por un empleo formal, mientras que este último estado civil genera un efecto positivo para ser seleccionado por los empleadores formales.

Al igual que en la especificación univariante, la variable de situación de desempleo anterior al actual empleo no se relaciona positivamente con juntarse a la cola. Este resultado puede implicar que una fase de desempleo anterior implica una mayor presión por encontrar algún tipo de empleo, por lo que los individuos prefieren ubicarse rápidamente en un puesto informal y no esperar el proceso para un empleo formal. De igual forma, se ha obtenido que los trabajadores del comercio, hostelería y construcción tienen una mayor probabilidad de ser informal ya sea porque no se han juntado a la cola de empleo o porque no han sido seleccionados de ella.

Respecto a las variables que sólo se incluyen en la ecuación de “juntarse a la cola” los resultados no difieren mucho de los obtenidos en la especificación univariante. Los niños en el hogar en edades extremas presentan un efecto negativo y un mayor nivel educativo en el hogar, otros miembros del hogar en la formalidad o en el desempleo generan un impacto positivo para localizarse en la formalidad.

De la estimación simultánea se puede observar que la correlación entre las dos ecuaciones (JC y EDC) es negativa (-0.58) y estadísticamente significativa. Este resultado indica que el proceso de localización laboral entre la formalidad e informalidad se lleva a cabo más de manera simultánea que en forma secuencial y existen factores inobservables que afectan dicho proceso.

5.3. Ecuaciones salariales para formales e informales

Las estimaciones de las ecuaciones salariales para formales e informales se muestran en el cuadro 7. Se observa que con excepción del parámetro σ_{2v_1} , los coeficientes estimados de la corrección de selectividad son estadísticamente significativos, además las pruebas conjuntas de estos parámetros para determinar si el sesgo de selección no es significativo han sido rechazadas. La principal implicación del sesgo de selección es que las tasas de retornos al capital humano en general y a la educación en particular son sobreestimadas con la estimación MCO, por ello la importancia de corregirla cuando existe tal sesgo (Heckman 1979).

Cuadro 7. Ecuaciones salariales para formales e informales corregidas con el probit bivariante
(Y= Log salario real mensual hora)

	MCO sin corrección				Corrección con el modelo Abowd-Farber				Corrección con el modelo Poirier	
	Formales		Informales		Formales		Informales		Formales	
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Constante	7.350***	186.51	-4.871***	-231.38	-4.058***	-81.16	-4.315***	-50.0	-4.075***	-82.01
Estudios primarios	-0.038	-1.00	0.189***	10.62	-0.054	-1.39	0.165***	8.67	-0.051	-1.31
Estudios secundarios	0.135***	3.50	0.419***	22.66	0.101**	2.51	0.284***	12.88	0.108***	2.71
Estudios superiores	0.782***	20.03	1.036***	50.28	0.726***	17.13	0.663***	24.25	0.737***	17.48
Experiencia	0.017***	19.84	0.019***	25.56	0.016***	18.5	0.011***	14.6	0.016***	18.75
Experiencia ²	-0.0003***	-19.66	-0.0003***	-29.7	-0.0003***	-17.85	-0.0002***	-14.13	-0.0003***	-18.22
Antigüedad	0.019***	43.97	0.006***	5.55	0.019***	41.62	0.009***	18.82	0.019***	42.41
Sexo (hombre=1)	0.068***	12.12	0.245***	34.82	0.057***	3.32	-0.132***	2.37	0.056***	2.67
Jefe de hogar (jefe de hogar=1)	0.095***	16.3	0.179***	28.71	0.093***	15.17	0.156***	23.35	0.094***	15.01
Casado (casado=1)	0.051***	9.36	0.076***	12.95	0.051***	8.91	0.095***	12.64	0.051***	8.67
Industria, electricidad, gas y agua	0.054***	7.67	-0.049***	-5.65	0.040***	4.00	-0.081***	-4.70	0.043	4.51
Transporte y comunicación	-0.046***	-4.47	-0.211***	-23.18	-0.053***	-5.10	-0.262***	-27.72	-0.052***	-4.93
Seguros, finan., admon pública, educ.	0.266***	36.72	0.196***	18.98	0.244***	22.52	0.073***	3.98	0.248***	23.16
Serv. personales, domésticos u otros	0.115***	13.06	0.092***	11.25	0.106***	10.24	0.089***	5.44	0.108***	10.73
σ_{1v_1}					-0.050***	-3.47			-0.045***	-2.93
σ_{2v_1}					-0.018	-0.66			-0.020**	-0.52
σ_{1v_2}							-0.408***	-22.74		
δ_1							-0.132***	-6.79		
δ_2							0.104***	3.31		
N	53383		87669		53383		87659		53383	
R2 ajust	0.36		0.16		0.36		0.17		0.36	

Fuente: Cálculos propios a partir de la GEIH 2009

*** p<0.01, ** p<0.05, * p<0.1

Errores estándar robustos a la heteroscedasticidad

Categorías base: Educación: Sin estudios; Sector: comercio, hostelería y construcción.

El Log salario real mensual hora está medido en cientos de miles de pesos colombianos de 2008.

Las estimaciones corregidas con el modelo secuencial (modelo Abowd-Farber) muestran que las variables de capital humano son factores positivos importantes en la determinación de los salarios, tanto en el sector formal como en el informal. La educación, sobre todo los estudios superiores, generan el mayor efecto. En los informales este efecto positivo puede indicar que los empleos de una parte de estos trabajadores no son secundarios y, por tanto, existen retornos por la educación (por ejemplo, para

el 17% de los informales que tienen estudios superiores (cuadro 2)). Para las características personales, se observa que existen diferencias salariales por sexo en ambos sectores, los jefes de hogar y los casados presentan una prima salarial respecto a los no jefes y a los solteros, y esta prima es mayor en el sector informal.

Con el modelo simultáneo (modelo Poirier) sólo fue posible hacer la corrección para la ecuación salarial de los formales (véase sección 3). Se observa que no existen marcadas diferencias respecto a la corrección con el modelo secuencial. Los coeficientes de selectividad son igualmente negativos y significativos, hay retornos positivos de la educación, la experiencia también presentan retornos positivos pero cóncavos, al igual que la antigüedad y existen primas salariales positivas por ser hombre, jefe de hogar y estar casado..

Las anteriores estimaciones además de ser usadas para corroborar la existencia de la cola por empleos formales (siguiente sección), se utilizan para calcular el diferencial salarial entre el sector formal y el informal, el cual se incorpora como regresor adicional en el probit bivariante estructural (sección 5.5). Analicemos entonces la existencia de la cola de empleo.

5.4. Pruebas de existencia de la cola de empleo

Para probar la validez de los resultados obtenidos con los modelos probit bivariantes y, por tanto, la existencia de una cola para los empleos formales se calculan los tests de Abowd y Farber, y Mengistae expuestos en la sección 3.2. Los primeros autores a partir de un test de razón de verosimilitud contrastan las hipótesis nula de no existencia de una cola de empleo y existencia de una cola universal a partir del modelo probit bivariante. La primera hipótesis implica que los coeficientes de la ecuación de “ser escogido de la cola” son cero con excepción de la constante, mientras que la segunda restringe los coeficientes de la ecuación de “juntarse a la cola” a cero. Con un test diferente, Mengistae propone contrastar la existencia o no de un racionamiento de puestos de trabajo formales a partir de pruebas sobre el parámetro π que aparece en las ecuaciones salariales condicionadas. La hipótesis nula de existencia de cola de empleo es $\pi = 0$ y la hipótesis nula de una cola universal es $\pi = 1$. La primera hipótesis se prueba haciendo un test de significancia conjunta de $\delta_1 = \delta_2 = 0$ y la segunda se prueba con un test F de $\sigma_{1v_1} = \delta_1$. El cálculo de las anteriores pruebas se muestra en el cuadro 8.

Los resultados de los test muestran que a un nivel de significancia de 5% se rechaza tanto la hipótesis nula de no existencia de una cola de empleo, como la existencia de una cola universal. Esto muestra evidencia a favor de la especificación bivariante frente a la univariante para describir el proceso de localización laboral entre la formalidad y la informalidad. Se tiene, entonces, que la existencia de una cola de empleo

Cuadro 8. Test de existencia de una cola de empleo

Ho	Test de Abowd-Farber		Test de Mengistae	
	Modelo Abowd-Farber	Modelo Poirier	Modelo Abowd-Farber	Modelo Poirier
No hay cola	LR Chi2(11) = 915.02 Prob >Chi2 = 0.000	LR Chi2(11) = 972.73 Prob >Chi2 = 0.000	F(2, 87642) = 23.18 Prob >F = 0.000	-
Cola universal	LR Chi2(21) = 5895.34 Prob >Chi2 = 0.000	LR Chi2(21) = 5987.48 Prob >Chi2 = 0.000	F(1, 87642) = 172.55 Prob >F = 0.000	-

Fuente: Cálculos propios a partir de la GEIH 2009

Nota: El test de Mengistae no es posible aplicarlo en el modelo Poirier ya que no se puede deducir la ecuación salarial condicional para los informales (véase sección 3.1).

indica un racionamiento de los puestos de trabajo formales. En este sentido no todo el que quiera emplearse en la formalidad lo puede hacer, debido a una restricción que existe en la generación de empleos formales. Siguiendo a [Mengistae \(1999\)](#), a partir de la ecuación (18), se ha podido deducir que el valor estimado de la proporción de trabajadores informales que han sido racionados del sector formal es de 32 %. Por su parte, el rechazo de una cola universal indica que no todos los trabajadores buscan localizarse en el sector formal. Surgen, entonces, dos tipos de informales: los que obtienen beneficios o ganan más por encontrarse en la informalidad; y otro tipo de informales que dada su baja cualificación optan por este sector como única alternativa de subsistir y escape al desempleo.

5.5. Determinantes de la decisión de formalidad/informalidad: modelo estructural bivalente

El probit bivalente estructural se ha estimado incluyendo en la ecuación JC el diferencial salarial entre el sector formal e informal, y en la ecuación EDC el salario esperado en el sector formal, el como regresores adicionales y se han excluido las variables que se asumen sólo determinan los salarios (variables dummies de sector económico). En la práctica existen dos tipos de salarios estimados: los incondicionales y los condicionales. Los primeros representan los salarios estimados que no tienen en cuenta el tipo de empleo que tienen los trabajadores. Esto es, por tanto, sin los efectos del sesgo de selección en los salarios (razones inversas de Mills). Los salarios estimados condicionales, por su parte, se basan en la posición relativa de los salarios del trabajador una vez el sector de empleo ha sido determinado. Esta última estimación de salarios se diferencia de la estimación no condicional en que los efectos de selección son debido a diferencias en las tasas de retorno de las características no observadas de los trabajadores y no a los niveles de dichas características ([Gyourko & Tracy 1988](#), [Mengistae 1999](#)). Para nuestro caso se van a utilizar las estimaciones de los salarios no condicionales, ya que interesa determinar si existe una prima salarial por ubicarse en la formalidad.

Cuadro 9. Estimaciones del modelo Abowd-Farber estructural
($Y_j = 1$ formal, 0 informal; $j = JC, EDC$)

	JC		EDC	
	Coef.	z	Coef.	z
Constante	-1.259***	-6.41	0.323***	2.85
<i>Prima laboral del sector formal</i>				
$\ln W_f - \ln W_{inf}$	4.250***	31.1		
$\ln W_f$			0.353***	15.68
<i>Características personales</i>				
Estudios primarios	0.061	0.32	1.002***	10.61
Estudios secundarios	0.079	0.41	1.481***	16.29
Estudios superiores	-0.080	-0.40	1.579***	17.3
Experiencia	0.005***	3.04		
Experiencia ²	-0.0003***	-8.67		
Antigüedad	-0.020***	-11.09		
Sexo (hombre=1)	-0.363***	-8.47	-0.384***	-9.44
Jefe de hogar (jefe de hogar=1)	0.550***	19.93	0.098***	4.10
Casado (casado=1)	0.189***	7.56	0.065***	2.70
<i>Características del hogar</i>				
# niños 0-2 años	-0.131***	-7.02		
# niños 3-5 años	-0.155***	-9.31		
# niños 6-10 años	-0.122***	-10.52		
# niños 11-17 años	-0.199***	-21.61		
# de formales en el hogar	0.444***	26.61		
# desempleados en el hogar	0.050***	4.03		
Educación promedio del hogar	0.029***	8.97		
<i>Características del empleo</i>				
Desempleado antes del actual empleo	-0.454***	-18.12	-0.194***	-7.51
N		154613		
Log L		-82414.38		
AIC		164884.80		
BIC		165163.30		

Fuente: Cálculos propios a partir de la GEIH 2009

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Categorías base: Educación: menos de estudios primarios

JC: ecuación de "juntarse a la cola"; EDC: ecuación de "ser escogido de la cola"

Las estimaciones máximo verosímil del modelo probit estructural secuencial con observabilidad parcial se muestran en el cuadro 9. Se observa que el diferencial salarial entre el sector formal e informal es el factor más importante para la ubicación en un empleo formal. De este modo entre mayor sea la prima salarial del sector formal, es más probable que un individuo se junte a la cola por un empleo de este tipo. En cuanto a la variable de salario esperado en el sector formal ($\ln W_f$), se observa que presenta un signo positivo y es estadísticamente significativa. Esto indica que una mayor productividad esperada dadas las cualidades del individuo aumenta la probabilidad que el empleador lo seleccione de la cola para la vacante formal.

Respecto a las otras variables se observa que la educación no presenta ningún efecto en la determinación de la búsqueda de empleo. La prima salarial que ofrece el sector formal se lleva todo el efecto. Esto puede indicar que los individuos se fijan

principalmente en esta última variable para buscar empleo, independientemente del nivel educativo que tengan. Por su parte, en la ecuación de ser seleccionado de la cola, se nota que la educación es el principal determinante, por encima del salario esperado en el sector formal. Este resultado muestra que la educación es el principal mecanismo que tienen los empleadores para determinar las cualidades de un trabajador y que le permite hacer un escalafón para seleccionar el mejor que haya para el empleo.

En cuanto a las otras características personales, se observa que ser hombre disminuye la probabilidad tanto de buscar un empleo formal como de ser seleccionado de la cola por los empleadores formales. Este resultado es consistente con el proceso de inserción laboral y mayores niveles educativos que ha tenido la mujer en los últimos años, además muestra la presión social y cultural que existe sobre los hombres para que se ubiquen rápidamente en un empleo, independientemente de su calidad (Uribe et al. 2006). La jefatura del hogar y estar casado presentando un efecto positivo en todo el proceso de localización en un empleo formal. Las mayores responsabilidades y el reflejo de un mayor compromiso en el empleo generan una mayor probabilidad de localizarse en la formalidad. Por último, se observa que una etapa de desempleo anterior tiene un efecto negativo tanto en la probabilidad de juntarse a la cola como de ser seleccionado por los empleadores. El comportamiento de las otras variables que sólo se incluyen en la ecuación de juntarse a la cola, permanecen igual que en la estimación reducida.

A partir de las anteriores estimaciones es posible estimar diferentes probabilidades. En particular se calculan las siguientes probabilidades:

- Prob. de juntarse a la cola (P(JC)): $Prob(V_{1i}^* > 0) = \Phi(X'_{1i}\hat{\beta}_1)$
- Prob. de ser escogido de la cola (P(EDC)): $Prob(V_{2i}^* > 0 | V_{1i}^* > 0) = \Phi(X'_{2i}\hat{\beta}_2)$
- Prob. de ser formal (P(formal)): $Prob(V_{1i}^* > 0 \text{ y } V_{2i}^* > 0) = \Phi(X'_{1i}\hat{\beta}_1)\Phi(X'_{2i}\hat{\beta}_2)$

Siguiendo a Farber (1983) y Venti (1987) es posible deducir la longitud de la cola. Ésta se calcula como la inversa del promedio de la probabilidad de ser escogido de la cola dado que el trabajador se ha juntado a la cola, esto es:

$$\text{Longitud de la Cola} = \frac{1}{\frac{1}{n} \sum_{i=1}^n P(EDC)}$$

En el cuadro 10 se muestra el resultado de las medias de las anteriores probabilidades y la longitud de la cola para diferentes características. Un primer aspecto interesante que se observa es la baja probabilidad que existe de ser formal (36 %), sólo los individuos con estudios superiores presentan una alta probabilidad de ubicarse en la formalidad (65 %). Respecto a la longitud de la cola se tiene que esta es de 1.62,

Cuadro 10. Probabilidades y longitud de la cola a partir del modelo bivalente estructural

	Toda la muestra	Hombre	Mujer	Sin estudios	Estudios primarios	Estudios secundarios	Estudios superiores
P(JC)	0.56	0.63	0.47	0.44	0.33	0.50	0.85
P(EDC)	0.61	0.56	0.68	0.13	0.44	0.64	0.76
P(formal)	0.36	0.38	0.34	0.06	0.14	0.32	0.65
Longitud de la cola	1.62	1.78	1.47	7.89	2.27	1.57	1.31

	Jefe de hogar	No jefe de hogar	Soltero	Casado	Ocupado antes del actual empleo	Desempleado antes del actual empleo
P(JC)	0.59	0.53	0.54	0.56	0.59	0.44
P(EDC)	0.61	0.62	0.60	0.62	0.62	0.57
P(formal)	0.38	0.34	0.35	0.37	0.39	0.27
Longitud de la cola	1.63	1.62	1.65	1.60	1.58	1.74

Fuente: Cálculos propios a partir de la GEIH 2009.

JC: ecuación de "juntarse a la cola"; EDC: ecuación de "ser escogido de la cola"

es decir que por cada 100 trabajadores en el sector formal, hay otros 162 que desean obtener un empleo formal. Este valor es considerablemente alto en comparación a la longitud de la cola de empleo formal que hay en Brasil (1.19) y Chile (1.06) (Soares 2004, Puentes & Contreras 2009).

Un aspecto importante del análisis de las probabilidades es ver como estas varían para diferentes características. Se observa, por ejemplo, que las mujeres, los individuos con más bajo nivel educativo y los que estuvieron en una etapa de desempleo anterior al actual empleo, son el grupo de trabajadores con menores probabilidades de buscar un empleo formal. En cuanto a la probabilidad de recibir una oferta de empleo formal, se tiene que las mujeres, los individuos con educación superior, los casados y con una ocupación previa al actual empleo, son los que mayor probabilidad presentan. Es de resaltar las diferencias en esta última probabilidad entre hombres (56 %) y mujeres (58 %). Nuevamente la mayor inserción de la mujer en el mercado laboral y el mayor nivel educativo en comparación con el hombre, que se ha presentando en la última década (Badel & Peña 2010, Hoyos et al. 2010), ejercen un efecto positivo en la decisión de contratación de los empleadores.

Respecto a la longitud de la cola, se tiene que los trabajadores sin estudios deben enfrentar a una cola de 7.89. De igual forma, los hombres y los solteros deben enfrentarse a una cola larga para un empleo formal: por cada trabajador hombre y soltero en el sector formal hay 1.78 y 1.65 trabajadores, respectivamente, que desean un empleo formal. También alarga la cola una fase de desempleo anterior, esta es de 1.74.

En síntesis, las estimaciones corroboran la existencia de un racionamiento de empleos formales. Tal racionamiento es reflejo de la poca capacidad del sector moderno o formal de la economía para generar los suficientes puestos de trabajo, surgiendo así

una cola de empleo y mecanismos por parte de los empleadores para ocupar las pocas vacantes. Ante estos dos mecanismos de selección que existe en el mercado laboral, la ubicación en un mejor empleo depende en gran parte de la prima salarial que ofrece el sector formal y las características de los individuos (sobre todo el capital humano), las cuales potencian la búsqueda e incrementan el número de ofertas de empleo que puede recibir el trabajador.

La prima salarial, junto con el nivel educativo, el compromiso y la responsabilidad en el empleo, son los principales factores que determinan que un trabajador se ubique en la formalidad. De igual forma, las variables relacionadas con el hogar, juegan un papel determinante en la busque de empleo formal/informal. La presión económica de los hijos menores en el hogar, ejercen un efecto negativo para localizarse en la formalidad. La necesidad de recursos aumenta el costo de oportunidad de la búsqueda de un empleo formal, por ello los individuos prefieren generar ingresos con mayor rapidez y menores trabas legales. En contra, un mejor clima educativo y mayor grado de formalidad en el hogar ejercen efectos positivos para que los individuos se ubiquen en la formalidad.

6. Conclusiones

Con una metodología alternativa a las comúnmente utilizadas en el análisis de los factores que influyen las decisiones de formalizarse/informalizarse en Colombia, este trabajo ha mostrado evidencia del racionamiento de empleos formales que existe en el mercado laboral colombiano. La poca generación de empleo formal producto de la estrechez del sector moderno, lleva a que se genere un doble mecanismo de selección para ubicarse en la formalidad/informalidad. En primer lugar, los individuos deben enfrentarse a una cola por los empleos formales. Juntarse a esta cola depende principalmente de la prima salarial que le ofrece el sector formal y el nivel de responsabilidad que tiene el individuo en el hogar.

En una segunda etapa, una vez el individuo se encuentra en la cola por un empleo formal, debe esperar por una oferta de empleo. Esta oferta de empleo depende principalmente el nivel de cualificación, el grado de responsabilidad, el compromiso y el historial laboral del individuo. Así, trabajadores con alto nivel educativo, con un mayor grado esperado de responsabilidad y buen historial laboral incrementan la probabilidad de ser seleccionado por un empleo formal.

En términos cuantitativos se ha podido corroborar que un 32% de los trabajadores que se localizan en la informalidad son producto del racionamiento de empleos formales. Este resultado muestra que existen dos tipos de informales: unos que no tienen otra alternativa más que la informalidad como medio de empleo, y otros que se ben-

efician de ubicarse en este sector. Esta última población ve a la informalidad como sector donde es posible acumular y en donde hay retornos por la educación, es decir que la informalidad deja de ser un sector secundario como es para los primeros. Sin embargo, debe estudiarse con mayor profundidad las características de este segundo tipo de informales, ya que se pueden estar incurriendo en sesgos por no tener en cuenta las preferencias de los trabajadores por cambiar de un sector a otro. Esta limitación es propia de la encuesta utilizada⁸.

Finalmente, la corroboración de un doble mecanismo de selección en el mercado laboral para ubicarse entre la formalidad y la informalidad, es evidencia del proceso de segmentación laboral que existe en Colombia. La escasez de empleos formales junto con una mala calidad de los existentes son las principales características. Son estos dos factores en los que deben centrarse las políticas.

Referencias

- Abowd, J. & Farber, H. (1982), 'Job Queue and the Union Status of Workers', *Industrial and Labor Relations Review*, **35**(3), 354–367.
- Archibald, R. B. (1977), 'Labor Queues and Involuntary Unemployment', *Economic Inquiry*, **15**(1), 33–55.
- Badel, A. & Peña, X. (2010), 'Decomposing the Gender Wage Gap with Sample Selection Adjustment Evidence from Colombia', *Revista de Análisis Económico*, **25**(2), 169–291.
- Bernal, R. (2009), 'The Informal Labor Market in Colombia: Identification and Characterization', *Desarrollo y Sociedad*, **63**, 145–208.
- Bosch, M. & Esteban-Pretel, J. (2009), 'The Informal Labor Market in Latin America', mimeo, The University of Tokyo.
- Bosch, M. & Maloney, W. (2008), 'Cyclical Movements in Unemployment and Informality in Developing Countries', *Policy Research Working Paper No 4648*, The World Bank, Junio.
- Chong, A., Galdo, J. & Saavedra, J. (2007), 'Informality and Productivity in the Labor Market: Peru 1986-2001', *Working Paper No 609*, Research Department, BID, Julio.
- Co, C., Gang, I. & Yun, M. (2005), 'Self-Employment and Wage Earning in Hungary', *Review of Development Economics*, **9**(2), 150–165.

⁸ Para el caso de Brasil, Soares (2004) puede distinguir entre los trabajadores informales que desean cambiar su empleo actual por uno formal y amplía el estudio teniendo en cuenta el sesgo de selección que implica este deseo por un empleo formal.

- Contreras, D., Mello, L. D. & Puentes, E. (2008), 'Tackling Business and Labour Informality in Chile', *OECD Economics Department Working Papers* No 607, OECD Publishing, Abril.
- De Soto, H. (1987), *El Otro Sendero. La Revolución Informal*, Instituto Libertad y Democracia, Lima.
- Dickens, W. & Lang, K. (1985), 'A Test of Dual Labor Market Theory', *The American Economic Review*, **75**(4), 792–805.
- Farber, H. (1983), 'The Determination of the Union Status of Workers', *Econometrica*, **51**(5), 1417–1437.
- Galli, R. & Kucera, D. (2008), 'Gender, Informality and Employment Adjustment in Latin America', *Working Paper No 85*, ILO, Policy Integration and Statistics Department - Geneva.
- Gyourko, J. & Tracy, J. (1988), 'An analysis of public and private sector wages allowing for endogenous choices of both government and union status', *Journal of Labor Economics*, **6**, 229–253.
- Heckman, J. (1976), 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simpler Estimator for Such Models', *Annals of Economic and Social Measurement*, **5**(4), 475–492.
- Heckman, J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica*, **47**(1), 153–161.
- Hoyos, A., Ñopo, H. & Peña, X. (2010), 'The Persistent Gender Earnings Gap in Colombia, 1994-2006', *IZA Discussion Paper* No 5073, Institute for the Study of Labor (IZA), Julio.
- Huguet, A. (1996), 'Dualidad en el Mercado de Trabajo Español', *Revista de Economía Aplicada*, **4**(11), 81–104.
- Jütting, J. & De Laiglesia, J. R. (2009), *Is Informal Normal? Towards More and Better Jobs in Developing Countries*, Technical report, OECD Development Center. 163 pág.
- Khamis, M. (2009), 'A Note on Informality in the Labor Market', *IZA Discussion Paper* No 4676, Institute for the Study of Labor (IZA), Diciembre.
- Lang, K. & Dickens, W. (1987), 'Neoclassical and Sociological Perspectives on Segmented Labor Market', *NBER Working Paper Series*, No 2127, NBER.
- Lee, L. F. (1978), 'Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables', *International Economic Review*, **19**(2), 415–433.

- Lehmann, H. & Pignatti, N. (2007), 'Informal Employment Relationships and Labor Market Segmentation in Transition Economies: Evidence from Ukraine', *IZA Discussion Paper No 3269*, Institute for the Study of Labor (IZA), Diciembre.
- Maddala, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.
- Maddala, G. S. & Nelson, F. (1975), 'Switching Regression Models with Exogenous and Endogenous Switching', *Proceedings of the American Statistical Association (Business and Economics Section)*, pp. 423–426.
- Maloney, W. (1998), 'The Structure of Labor Markets in Developing Countries. Time Series Evidence on Competing Views', *Policy Research Working Paper No 1940*, The World Bank.
- Maloney, W. (2004), 'Informality Revisited', *World Development*, **32**(7), 1159–1178.
- Meng, C. & Schmidt, P. (1985), 'On the Cost of Partial Observability in the Bivariate Probit Model', *International Economic Review*, **26**(1), 71–85.
- Mengistae, T. (1999), 'Wage Rates and Job Queue: Does the Public Sector Overpay in Ethiopia', *World Bank Policy Research Working Paper No 2105*, The World Bank, Abril.
- Mohanty, M. (2001), 'Testing for the Specification of the Wage Equation: Double Selection Approach or Single Selection Approach', *Applied Economics Letter*, **8**, 525–529.
- Mondragón-Vélez, C., Peña, X. & Willis, D. (2010), 'Labor Market Rigidities and Informality in Colombia', *Documento CEDE No 2010-07*, Universidad de los Andes, Bogotá, Febrero.
- Núñez, J. (2002), 'Empleo Informal y Evasión Fiscal en Colombia', *Archivos de Economía No 210*, DNP.
- OIT (2009a), *Panorama Laboral 2009. América Latina y el Caribe*. OIT, Ginebra.
- OIT (2009b), *Tendencias Mundiales del Empleo 2009*. OIT, Ginebra.
- Ortiz, C., Uribe, J. & Vivas, H. (2010), 'Para Generar Empleo Formal Hay que Crear Empresas Modernas', *Documento de Trabajo No 129*, CIDSE, Universidad del Valle, Mayo.
- Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A. & Saavedra-Chanduvi, J. (2007), *Informality: Exit and Exclusion*, The World Bank.
- Poirier, D. J. (1980), 'Partial Observability in Bivariate Probit Models', *Journal of Econometrics*, **12**, 209–217.
- Portes, A. (1995), *En Torno a la Informalidad: Ensayos Sobre Teoría y Medición de la Economía no Regulada*, Miguel A. Porrúa Editor, México.

- Puentes, E. & Contreras, D. (2009), 'Informal Jobs and Contribution to Social Security: Evidence from a Double Selection Model', *Serie Documentos de Trabajo No 307*, Universidad de Chile, Diciembre.
- Ribero, R. (2003), 'Gender Dimensions of Non-Formal Employment in Colombia', *Documento CEDE No 2003-04*, Universidad de los Andes, Febrero.
- Roldan, P. & Ospina, C. (2009), '¿Quiénes Terminan en la Informalidad?: Impacto de las Características y el Tiempo de Búsqueda', *Revista de Economía del Caribe*, **4**, 149–180.
- Saavedra, J. & Chong, A. (1999), 'Structural Reform, Institutions and Earnings: Evidence from the Formal and Informal Sectors in Urban Peru', *Journal of Development Studies*, **35**(4), 95–116.
- Schneider, F. & Klinglmair, R. (2004), 'Shadow Economies Around the World: What do we Know', *Working Paper No 2004-03*, CREMA.
- Soares, F. (2004), 'Do Informal Workers Queue for Formal Jobs in Brazil?', *Working Paper No 1021*, IPEA, Mayo.
- Tunali, I. (1986), "A General Structure for Models of Double-Selection and an Application to Joint Migration/Earning Process with Remigration", in R. D. Ehrenberg, ed., 'Research in Labor Economics', Vol. 8, Part B. Greenwich. MA.: JAI Press, pp. 235-282.
- Uribe, J. & Ortiz, C. (2006), *Informalidad Laboral en Colombia 1988-2000: Evolución, Teorías y Modelos*, Programa Editorial de la Universidad del Valle. Cali-Colombia.
- Uribe, J., Ortiz, C. & Correa, J. (2006), '¿Cómo Deciden los Individuos en el Mercado Laboral?', *Lecturas de Economía*, **64**, 59–89.
- Uribe, J., Ortiz, C. & García, G. (2007), 'La Segmentación del Mercado Laboral Colombiano en la Década de los Noventa', *Revista de Economía Institucional*, **9**(16), 189–221.
- Venti, S. F. (1987), "Wages in the Federal and Private Sectors", in D. A. Wise, ed., 'Public Sector Payrolls', Chicago: University of Chicago Press, pp. 147-182.