

Cristian Darío Castillo Robayo*
Javier García Estévez**

Desempleo juvenil en Colombia: ¿la educación importa?

Recibido: 17 de agosto de 2018

Concepto de evaluación: 5 de diciembre de 2018

Aprobado: 8 de febrero de 2019

Artículo de investigación

© 2019 Universidad Católica de Colombia.

Facultad de Ciencias
Económicas y Administrativas.
Todos los derechos reservados

RESUMEN

Los jóvenes enfrentan obstáculos especiales en inserción productiva. Indicadores laborales como el desempleo, el empleo precario o los ingresos son significativamente peores en ellos en comparación con los de los adultos. Uno de los factores principales del desempleo juvenil es el desajuste de las competencias: la sobreeducación y el exceso de competencias coexisten con la subeducación y la escasez de competencias. Este artículo analiza los factores que inciden en la tasa de desempleo juvenil en Colombia, con especial atención al papel de la educación, pues varios hallazgos confirman la importancia de esta en la consecución de empleos formales, pero no en la probabilidad de estar ocupado.

Palabras clave: capital humano, desempleo juvenil, educación, empleo, informalidad, segmentación laboral.

JEL: J19, J23, J24.

Cómo citar este artículo / To reference this article / Para citar este artículo:

Castillo Robayo, C. D., & García Estévez, J. (2019). Desempleo juvenil en Colombia ¿la educación importa?. *Revista Finanzas y Política Económica*, 11(1), 101-127. doi: <http://doi.org/10.14718/revfinanzpolitecon.2019.11.1.7>

Youth unemployment in Colombia: Does education matter?

ABSTRACT

Young people face unique obstacles in their search for a productive work life, and more frequently than adults. Employment indicators for young people—such as unemployment rate, precarious employment, income, among others—are significantly worse than those of adults. One of the main factors of youth unemployment is the mismatch of competencies: over-education and over-skilling coexist with under-education and lack of skills. This article analyzes the factors that affect youth unemployment rate in Colombia. Particular attention is paid to the role of education. The main findings confirm the importance of education in achieving better quality jobs in terms of labor formality, but not in the probability of being employed.

* Doctor en economía (c). Universidad Autónoma de Madrid, España.
Correo: cristian.castillo@estudiante.uam.es
 <https://orcid.org/0000-0001-7569-1860>

** Doctor en economía. Universidad de los Andes. Bogotá, Colombia.
Correo: jh.garcia@uniaandes.edu.co
 <https://orcid.org/0000-0002-5796-245X>

Keywords: Youth unemployment, education and employment, informality.

Desemprego juvenil na Colômbia: a educação importa?

RESUMO

Os jovens enfrentam obstáculos especiais para entrar no mundo do trabalho. Indicadores de trabalho como desemprego, emprego precário ou ingressos são significativamente piores neles em comparação com os dos adultos. Um dos fatores principais do desemprego juvenil é o desajuste das competências: a “supereducação” e o excesso de competências coexistem com a “subeducação” e a escassez de competências. Este artigo analisa os fatores que incidem na taxa de desemprego juvenil na Colômbia, com ênfase no papel da educação, pois vários achados confirmam a importância desta para conseguir empregos formais, mas não para a probabilidade de estar ocupado.

Palavras-chave: capital humano, desemprego juvenil, educação, emprego, informalidade, segmentação no trabalho.

INTRODUCCIÓN

El desempleo juvenil se ha convertido en un problema de escala mundial debido a sus costos económicos y sociales y a la magnitud de sus efectos en el desarrollo de largo plazo en las sociedades. En los últimos años, el debilitamiento de la recuperación económica mundial ha agravado la crisis del empleo juvenil, dificultando aún más el acceso de muchos jóvenes al trabajo. Uno de los principales factores del desempleo juvenil es el desajuste de las competencias: la sobreeducación y el exceso de competencias coexisten con la subeducación y la escasez de estas, y cada vez más con el desgaste de la formación adquirida por causa del desempleo de larga duración (Organización Internacional del Trabajo [OIT], 2013).

Uno de los enfoques más utilizados para estudiar el desempleo juvenil es el análisis de la transición de la escuela al trabajo y, por tanto, el análisis de la interacción entre educación y mercado de trabajo. La teoría del capital humano estudia la conexión entre mayores cualificaciones provenientes de la educación y una mayor productividad (Schultz, 1961; Becker, 1964). En concreto, Schultz (1961) postuló la relación positiva entre mayores niveles de educación y mejores empleos y salarios, mientras Becker (1964) mostró la relación directa entre el crecimiento económico y un mayor porcentaje de población con niveles de educación altos. Aquí la educación no se ve como un gasto sino como una inversión que reporta rendimientos hacia futuro, vía mayor empleabilidad y salarios más altos. Con esta expectativa, los jóvenes, sus familias y el conjunto de la sociedad invierten parte de sus activos y patrimonio en mayores niveles educativos para obtener por esta vía mayores posibilidades de ocupación y mejores salarios. Sin embargo, como se analizará en este artículo, los datos sobre desempleo juvenil evidencian que la transición de la escuela al trabajo no se está dando de la forma deseada.

Como se ha visto, la educación desempeña un papel importante en la determinación del empleo y su calidad. El desempleo juvenil viene acompañado por el cambio generacional en el mundo laboral.

Al respecto, la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2008) muestra el aumento sustancial de jóvenes que ingresan al mercado de trabajo y cómo estos cada vez ocupan una proporción mayor de la fuerza laboral. Con ello surge la preocupación por la inversión en capital humano realizada en los últimos años; la literatura refiere que, a pesar de los niveles educativos, los jóvenes presentan altos niveles de desempleo y periodos prolongados de paro, fenómeno que se entiende como sobreeducación o degradación de empleo (Léné, 2011).

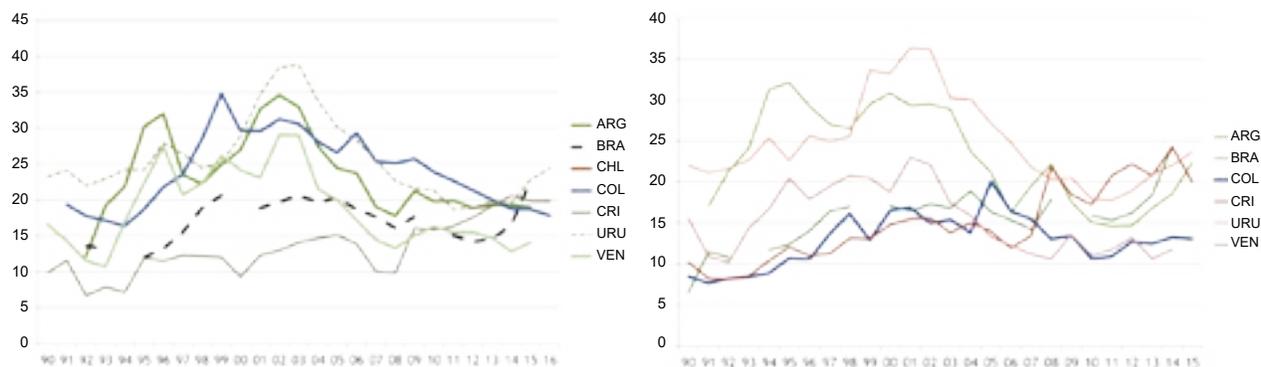
Este artículo se propone analizar los factores que inciden en la tasa de desempleo juvenil en Colombia, y en particular el papel de la educación en este fenómeno. El mercado de trabajo juvenil colombiano tiene particularidades que lo convierten en un interesante caso de estudio: en primer lugar, Colombia presenta la más alta tasa de desempleo de jóvenes con educación superior en América Latina (figura 1); en segundo lugar, durante varios años la tasa de desempleo de jóvenes colombianos ha doblado la tasa para el total de población en Colombia (figura 2).

La figura 1 revela el comportamiento de la tasa de desempleo juvenil por nivel de educación (alto y bajo)¹ en una muestra de países latinoamericanos. En cuanto a la evolución del desempleo juvenil de alta educación, Colombia aparece dentro de los países con mayor desempleo de jóvenes, superado solo por Argentina y Uruguay entre el 2001 y el 2006. La figura también muestra un rasgo positivo: el desempleo juvenil de alta educación ha disminuido en los últimos años. En lo relativo a las tasas de desempleo de los jóvenes con baja educación, Colombia tiene las cifras más bajas de la región, similares a las de Venezuela, Brasil y Costa Rica. Se hace evidente, en consecuencia, la brecha entre el desempleo de jóvenes con alta y baja educación en Colombia sin parangón en otro país latinoamericano.

1 Baja educación: 0-8 años de educación; alta educación: 14 o más años de educación.

Figura 1.

Tasa de desempleo juvenil por nivel educativo en América Latina



Fuente: elaboración propia a partir de datos del BID.

El *stock* de capital humano en Colombia ha aumentado de forma sostenida en las últimas décadas. Los jóvenes son quienes más capital humano aportan por rango de edad, y visto por género, son los hombres (adultos y jóvenes) los que más lo hacen (Correa y Montoya, 2013). En simultáneo aparece el hecho de que más trabajadores con educación superior entran al mercado de trabajo desplazando a los menos educados (Lasso, 2011). Sin embargo, otros trabajos como el de Gómez y Zárate (2011) establecen, para el caso de Argentina, Brasil y Colombia, una relación positiva entre aumento de inversión en educación y desempleo, más allá del efecto de la educación en los retornos salariales (como es la perspectiva de Castillo, Da Silva y Pérez, 2017).

REVISIÓN DE LA LITERATURA

En este apartado se revisan los estudios sobre desempleo en general y los análisis particulares para la población joven tanto en Colombia como a nivel internacional, para así poner en contexto la discusión sobre las causas y efectos de esta problemática.

Estudios para el caso colombiano

Los principales estudios de desempleo general y juvenil en Colombia han adoptado los siguientes

enfoques: el estudio de los flujos de los trabajadores y el ciclo económico, la segmentación laboral, la búsqueda de trabajo y la teoría del capital humano desde la perspectiva de la oferta laboral.

Flujo de trabajadores y ciclo económico

Estos estudios se encargan de analizar la entrada y salida del desempleo. Dentro de sus principales hallazgos se destacan los siguientes: la creación de empleo es procíclica y la destrucción es contracíclica (Rivas, 2004); existe movilidad de trabajadores por tipo de ocupación y por tamaño de las empresas; en perspectiva de largo plazo, los jóvenes comienzan como asalariados y terminan como independientes; finalmente, el flujo de trabajadores depende del ciclo económico (López, 1996). En las últimas décadas se ha evidenciado el crecimiento casi nulo del empleo neto (la diferencia entre creación y destrucción de empleo) y la destrucción de los empleos menos productivos por parte de las empresas en época de recesión (Melo y Ballesteros, 2013). Asimismo, el mercado de trabajo tiene un ciclo de corto plazo que viene determinado por el ciclo de la economía; los trabajadores independientes y no cualificados se retiran del mercado cuando hay auge económico y vuelven a él en época de crisis (Lasso, 2011).

Segmentación laboral

Los estudios de segmentación laboral asumen la existencia de dos sectores en el mercado de trabajo: uno formal y otro informal. De esta forma, se ha evidenciado el carácter estructural y creciente de los problemas de informalidad, así como la coexistencia de un sector formal con trabajadores asalariados y un sector de no asalariados con trabajadores por cuenta propia que cada vez ocupan un lugar mayor en la estructura del mercado laboral del país (Cárdenas y Mejía, 2007; Mondragón, Peña y Wills, 2010; Prada, 2012).

Colombia tiene una alta tasa de empleo informal, es decir, trabajadores que no están protegidos por las normativas, con relaciones laborales precarias y de baja remuneración. Así también, hay una correlación negativa entre nivel educativo e informalidad (García, 2010; Galvis, 2012; Guataquí, García y Rodríguez, 2010; Sánchez, 2015). Por otra parte, los trabajadores pobres tienen menos escolaridad y se ubican mayormente en la informalidad, aunque trabajan menos horas; no obstante, entre el 2002 y el 2012 se redujo la probabilidad de que el trabajador fuera pobre por el factor *educación* (Sánchez, 2015).

Otro rasgo importante de la segmentación laboral es la mayor participación de los jóvenes y las mujeres en el sector informal. Los jóvenes comienzan su ciclo de vida laboral en el sector informal y se mueven hacia el formal cuando acumulan experiencia. Sin embargo, cuando ya han apropiado suficiente capital humano, retornan al sector informal, esta vez como trabajadores independientes o propietarios de pequeñas empresas o negocios (Flores, 2002). Asimismo, los estudios de segmentación laboral han evidenciado que el hecho de ser jefe de hogar, hombre y contar con experiencia tiene impacto positivo en la participación laboral, que la educación favorece la formalidad (Uribe, Ortiz y Correa, 2004) y genera mayores retornos en el sector formal (Núñez, 2002), y que existe una relación positiva entre los asalariados ocupados y los empleos de calidad (Mora y Ulloa, 2011). Estudios como estos se unen a los de informalidad, que

parten de un análisis de mercados segmentados y problemas institucionales, como es el caso de la investigación de Camargo y Mora (2011).

Búsqueda de empleo

Estos estudios analizan los canales de búsqueda de empleo de los trabajadores y las características socioeconómicas que hacen posible encontrar trabajo. Así, la duración en el desempleo depende de características personales, la acumulación de capital humano (efecto aspirante) y el acceso a información y redes de relaciones (efecto oportunidad). Por su parte, los canales de búsqueda se clasifican en: a) informal, con asimetría en la información y sin normas; b) informal moderado, donde hay un trato directo de los individuos con los empleadores, y c) canal formal, en el que los mecanismos de vinculación están regulados por normas y se cumplen requisitos concretos para emplearse (Viáfara y Uribe, 2009).

En el periodo 1986-2002, Martínez (2003) encuentra que las personas más educadas (secundaria o universidad) tienden a ser más vulnerables al desempleo, y que los más jóvenes educados y quienes están en la informalidad presentan más inestabilidad laboral. De igual forma, Viáfara y Uribe (2009) constatan que las personas con maestría y doctorado duran menos en el desempleo, similar a aquellas que han acumulado experiencia. En contraste, las personas sin educación son los que más duran. En general, la educación reduce el tiempo en el desempleo para quienes usan canales informales, mientras para los otros dos canales lo aumenta. Los autores atribuyen este efecto a que las personas que usan el canal informal vienen de familias con contactos y más dinero. Por otra parte, comparando educación y experiencia, la primera reduce el tiempo de desempleo en un 1%, mientras que la segunda lo hace en un 4%.

Para el caso de Cartagena, Del Río, Yáñez y Pérez (2012) evidenciaron que el género, la experiencia y el canal de búsqueda impactan la salida del empleo. Así también, que la escolaridad,

la experiencia, el estrato y el género del jefe del hogar determinan la salida del desempleo. En este caso, los canales informales son más eficientes, las mujeres duran más tiempo desempleadas y las personas con educación secundaria consiguen empleo más rápido (en los primeros seis meses) que aquellas con educación universitaria y primaria; desde ese momento la educación universitaria presenta un efecto positivo.

Análisis de la demanda de trabajo

Pocos estudios han analizado la dinámica de la demanda de trabajo en Colombia. Entre ellos se encuentra el de Cárdenas *et al.* (2015), que demuestra que hay un problema de *mismatch* en gran parte debido a la falta de información en el mercado de trabajo, lo cual podría solucionarse con los avances del *big data* para conectar oferta y demanda. Por su parte, Guataquí *et al.* (2014) demuestran que la demanda laboral en Colombia genera una dinámica que favorece los salarios y la ocupación de bachilleres, técnicos, tecnólogos y universitarios; sin embargo, aunque se demandan más bachilleres, los salarios son mayores a mayor nivel educativo. Por su parte, con una metodología experimental, Atanasio, Kugler y Meghir (2007) evalúan el programa Empleo en Acción² y encuentran una relación positiva entre el programa, la cantidad de horas de trabajo y el salario, lo cual afecta también el nivel de ingreso de las familias.

Estudios recientes a nivel internacional

Los estudios sobre desempleo juvenil han tenido variados abordajes y aproximaciones. Por ejemplo, han analizado el efecto del salario mínimo, el capital humano, la segmentación de los mercados, la búsqueda de empleo y los choques de los ciclos de la economía. Aunque en varios temas hay sólidos consensos, en otros la controversia está activa.

2 Programa gubernamental de transferencias monetarias que busca que los jóvenes colombianos más pobres sigan estudiando en educación superior.

En esta sección se presentan los hallazgos de estudios recientes sobre desempleo juvenil en mercados de trabajo de países desarrollados, con el propósito de identificar coincidencias y diferencias con el caso colombiano.

Tradicionalmente se ha pensado que el aumento del salario mínimo provoca un aumento en el desempleo y una sustitución de mano de obra cualificada por una menos cualificada pero más barata. Sin embargo, para el caso de los jóvenes este aumento tiene un efecto positivo en el empleo, ya que incrementa la participación laboral y la empleabilidad (Giuliano, 2013). Además, puede ser un factor positivo y vinculante para los jóvenes con bajos niveles de educación, que en época de crisis pueden caer más en desempleo por esta rigidez del mercado. Por el contrario, los jóvenes más educados conservan más fácilmente el empleo, pero se ven sujetos a aceptar trabajos de bajos salarios (Cockx y Ghirelli, 2016).

Asimismo, las políticas de subsidio al salario o al salario mínimo pueden mejorar las oportunidades de los jóvenes para elevar su capital humano mediante formación laboral (Mroz y Savage, 2006). Este efecto puede lograrse también con las políticas activas de empleo y de seguros al desempleo, que mejoran el tiempo de búsqueda de trabajo y en el caso de los más educados permite encontrar mejores puestos de trabajo, aunque retrasa su emparejamiento (*match*) (Gaure, Røed y Westlie, 2012). Esto último puede ocurrir por el salario de reserva³ de las personas con mayor capital humano, que prefieren esperar más tiempo en el desempleo para encontrar trabajos de mayor calidad.

Además de los comentados efectos en la duración del desempleo, el salario mínimo y las políticas de subsidio inciden en la calidad del empleo. Un síntoma de ello es que hay cada vez más trabajadores altamente cualificados en puestos

3 El salario de reserva muestra el nivel de indiferencia de una persona respecto a aceptar un salario o rechazarlo y seguir en la búsqueda de empleo. Esto proporciona un límite para el individuo para emplearse o seguir en el desempleo, por lo que a mayor salario de reserva, menor será la posibilidad de que esta persona acepte las ofertas laborales y, por tanto, mayor la duración en el desempleo.

de trabajo que requieren bajos niveles educativos (McGuinness y Wooden, 2009), fenómeno atribuible a que las personas con mayores cualificaciones prefieren aceptar empleos por debajo de su nivel educativo mientras esperan uno de mejor calidad (Dolado, Jansen y Jimeno, 2009).

En una línea similar, Léné (2011) muestra que un incremento exógeno de trabajadores altamente calificados aumenta la proporción de estos en puestos de trabajo de baja calidad (*downgrading*), lo que además provoca que los trabajadores con bajo nivel educativo no se incorporen en el mercado, o bien, lo hagan en puestos de muy baja calidad⁴. El hecho de que los jóvenes acepten trabajos por debajo de sus cualificaciones puede empeorar su desempeño futuro. Baert, Cockx y Verhaest (2013) afirman que cuando hay sobreeducación⁵ se cae en una trampa, porque se afecta la tasa de transición de la persona a trabajos acordes con su educación. Este fenómeno se da especialmente en los jóvenes al inicio de su carrera (Carroll y Tani, 2013; Dolton y Vignoles, 2000) y afecta especialmente sus salarios (Montt, 2017; McGuinness, 2006)⁶.

El problema de la sobreeducación al inicio de la carrera conduce a tenerla también más adelante en el ciclo de vida laboral, lo cual podría evitarse mediante el mejoramiento de las aptitudes en el puesto de trabajo, en mayor medida para los jóvenes desplazados (Meroni y Vela, 2017). Sin embargo, las empresas no toman buenas decisiones de capacitación porque no valoran los beneficios de los trabajadores hacia futuro; además, los empresarios no valoran las externalidades por atracción de nuevos empleados ni la externalidad por desempleo (Belan y Chéron, 2014). Esto último,

aplicado al caso colombiano, se observa en Mora (2008), que señala que este no es un problema realmente grande (14% de sobreeducación) y tiene características particulares de oferta y demanda. Así, por ejemplo, la sobreeducación aumenta en las empresas pequeñas y en las que tienen mayor movilidad entre áreas de desempeño, y disminuye cuando se cuenta con mayor experiencia. En el salario además se castiga a los sobreeducados.

La acumulación de capital humano sigue siendo un factor importante para reducir el desempleo. Periodos prolongados de desempleo producen en el corto plazo inversiones no óptimas en capital humano entre los jóvenes, lo que afecta los salarios y el empleo en el largo plazo (Mroz y Savage, 2006). Adicionalmente, el capital humano específico no es transferible de un puesto de trabajo a otro, lo que restringe la movilidad. El retorno de la búsqueda de empleo de los adultos puede ser mayor a la de los jóvenes con mayor capital humano (Cheron, Hairault y Langot, 2013). Por consiguiente, los retornos del capital humano se reflejan principalmente en la madurez laboral y, dada su baja flexibilidad, pueden representar un problema ante choques económicos.

El análisis de la relación educación-empleo no arroja evidencia concluyente. Por una parte, autores como Lamo, Messina y Wasmer (2011) sustentan que la educación especializada reduce la movilidad de trabajadores y, por tanto, su capacidad para hacer frente a los cambios económicos. Así, en momentos de transición económica, tener título profesional genera periodos de desempleo más largos y tasas altas de desempleo para trabajadores de mayor edad, y esto es aún más evidente cuando no hay movilidad intersectorial. En contraste, Riddell y Song (2011) afirman que mayores niveles educativos mejoran los índices de reinserción al mercado de trabajo y mejoran la adaptación a los choques económicos, debido a la relación positiva entre educación, empleabilidad y salarios. Las recesiones económicas generan periodos de búsqueda de empleo más largos especialmente para los jóvenes menos educados y en mercados con mayor rigidez (Brunner y Kuhn, 2014).

4 Este autor argumenta además que en años anteriores las personas podían compensar con experiencia sus bajos niveles de educación, y esto les permitía acceder a mejores empleos; sin embargo, esto cambió, y la sola acumulación de experiencia ya no es suficiente.

5 Los autores entienden por sobreeducación el aceptar trabajos por debajo de las cualificaciones que tiene la persona.

6 Esto puede tomarse como un ejemplo de que las políticas activas de mercado de trabajo (PAMT) son útiles no solo en la empleabilidad sino en realizar un correcto *match*, ya que deben generarse políticas en pro de conectar el nivel educativo adecuado para no generar problemas en la transición laboral ni en los retornos del capital humano.

Otro punto de análisis del desempleo ha sido el estudio de redes: Pellizzari (2010) muestra que la familia y los amigos —los lazos débiles— son el mejor mecanismo para buscar empleo, ofreciendo menos costos y mayor confiabilidad; Zenou (2015) confirma que el aumento del tiempo dedicado a los lazos débiles aumenta la tasa de empleo.

METODOLOGÍA Y DATOS

Análisis descriptivo

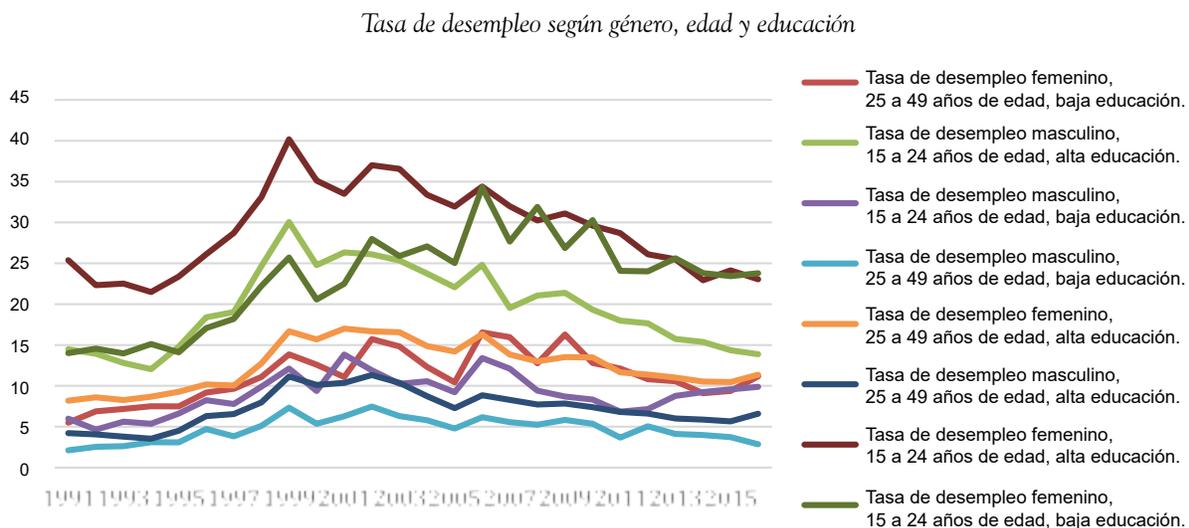
La figura 2 presenta la tasa de desempleo juvenil por género, edad y nivel educativo. Se encuentran en primer lugar las mujeres jóvenes y los grupos con alta y baja educación, aunque desde el 2006 hay una tendencia decreciente. A estos dos grupos les siguen hombres jóvenes con alta educación y mujeres adultas con alta y baja educación. En contraste, los hombres adultos con baja y alta educación y los hombres jóvenes con baja educación tienen las menores tasas de desempleo.

La figura 3 reporta el comportamiento de las vacantes generadas por los empresarios registradas en el Servicio Público de Empleo desde el 2014.

Se observa que los bachilleres, seguidos de los técnicos y universitarios, son los más demandados. En efecto, en el 2014 los bachilleres representaron el 23,5% de las vacantes demandadas por los empresarios, los técnicos representaron el 15,3% y los universitarios, el 23,1%. En el 2016 las vacantes demandaron en 33,7% bachilleres, 20,5% técnicos y 18,5% universitarios. En este periodo ha ocurrido un importante aumento de la demanda por bachilleres y una leve reducción de la demanda por universitarios.

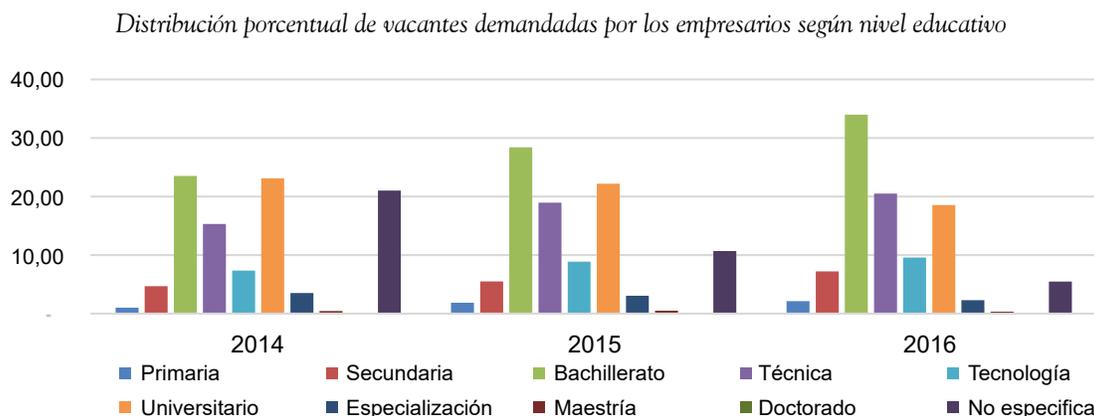
La figura 4 muestra el comportamiento de la informalidad por rango de edad y nivel educativo. Visto por rango de edad, en los mayores de 65 años se presentan los mayores porcentajes, algo esperable por ser pensionados o salir del mercado laboral. El siguiente grupo con mayor informalidad es el de los más jóvenes, con una tasa de informalidad que casi dobla la de los adultos. Vista por nivel educativo, la informalidad más alta se presenta en los menos educados y en los de educación media. Por el contrario, los más educados presentan la tasa de informalidad más baja. Así, tenemos indicios de la existencia de una relación inversa entre la probabilidad de estar ocupado y tener más nivel de educación, y de una relación directa entre estar más educado y estar en la formalidad.

Figura 2.



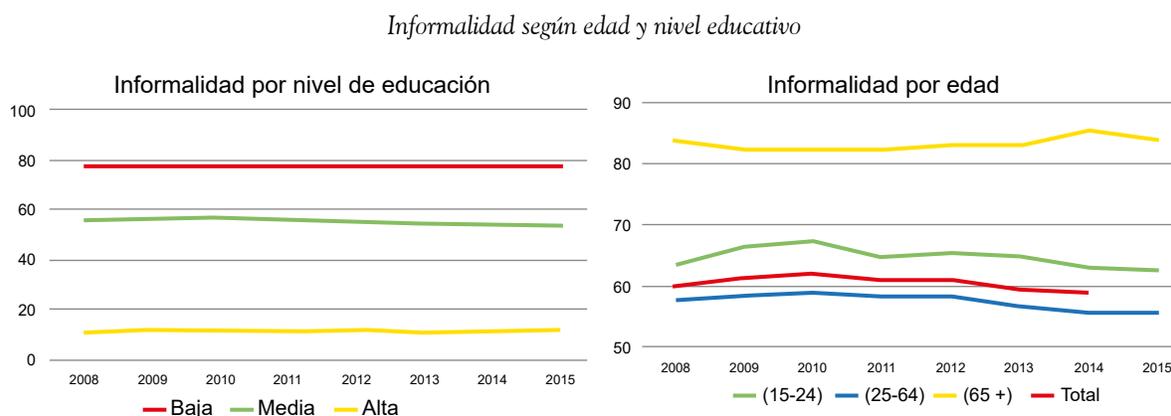
Fuente: elaboración propia a partir de datos del BID.

Figura 3.



Fuente: elaboración propia a partir de datos del Servicio Público de Empleo.

Figura 4.



Fuente: elaboración propia a partir de datos del DANE.

La figura 5 confirma los problemas que tienen los jóvenes en la informalidad. Según el tipo de ocupación y el sector de la economía, se puede ver que los obreros, el empleador particular y los trabajadores por cuenta propia presentan la mayor informalidad. Por sector económico, la mayor se presenta en el comercio, los restaurantes y hoteles, y en el sector de los servicios comunales, sociales y personales. Al comparar estas figuras con el tipo de ocupación de los jóvenes y el sector donde trabajan, se encuentra que los obreros, el empleador particular y los trabajadores por cuenta propia tienen las ocupaciones con mayor nivel de informalidad, seguidos por los sectores más informales: comercio,

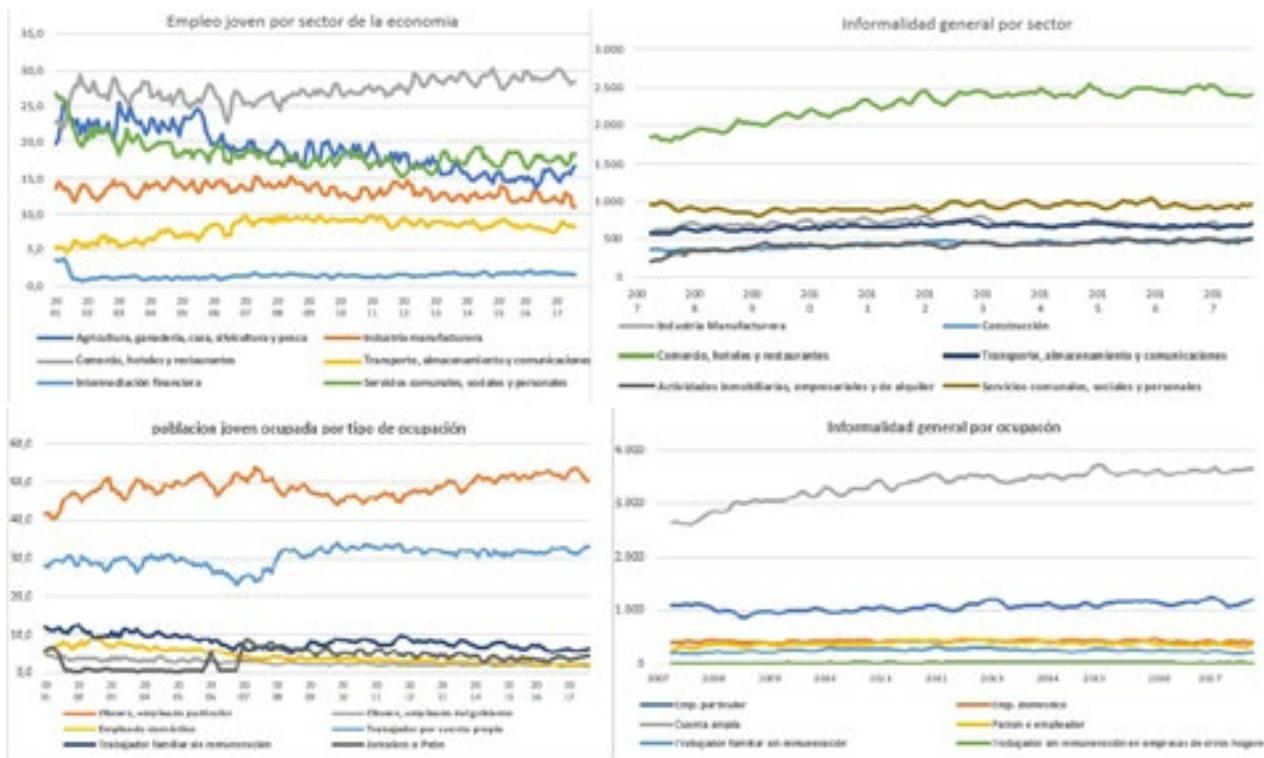
restaurantes y hoteles, y servicios comunales, sociales y personales.

Datos

El análisis empírico se realiza a partir de los datos de la Gran Encuesta Integrada de Hogares (GEIH) del 2016 aplicada por el Departamento Administrativo Nacional de Estadística (DANE). Los modelos por estimar usan tres variables dependientes. Las dos primeras son *empleo* y *empleo formal*; son variables *dummy* con valor 1 si se está ocupado y 0 en caso contrario. La tercera variable, que es usada en el modelo multilogit, está construida como una *variable*

Figura 5.

Informalidad según tipo de ocupación y sector económico



Fuente: elaboración propia a partir de datos del DANE.

de elección discreta con los siguientes valores: 1 = desocupado, 2 = empleo formal y 3 = empleo informal. Todos los valores vienen dados por la metodología del DANE, que determina la población inactiva y desocupada. Las variables de *empleo formal* e *informal* vienen de la población ocupada.

El modelo por estimar incluye cinco variables independientes: educación, género, región, jefe de hogar y unido. En primer lugar, la variable *educación* refiere el último título obtenido. La variable *género* captura el efecto de ser hombre o mujer; es una variable *dummy* con valores de hombre = 1 y mujer = 0. En tercer lugar, las variables *región*⁷ miden el efecto de estar en una zona geográfica y la probabilidad de estar ocupado. Por último, las variables

jefe de hogar y *unido*⁸ son variables categóricas que muestran si estas condiciones de las personas afectan la probabilidad de ocupación.

La variable *edad* se construye en grupos, en línea con la OIT (2015): *grupo edad1* son jóvenes entre los 16 y los 18 años; *grupo edad2* son jóvenes entre los 19 y 23 años; *grupo edad3* son jóvenes entre los 24 y 29 años (todos los anteriores marcan el periodo de juventud entre los 16 y 29 años y sus subgrupos); el *grupo edad4* son personas entre los 30 y 34 años, donde comienza la madurez laboral; finalmente, el *grupo edad5* está

7 Esta variable es una modificación de la de departamentos, con base en la clasificación del DANE.

8 Esta variable se basa en el estado civil de las personas; *está unido* (1) representa a los casados y en unión libre, y quienes no lo están (0) representan a los unidos pero no en unión libre (menos de dos años), viudos, solteros o divorciados.

Tabla 1.

Datos de las variables de los modelos (en porcentajes)

Variable	Inactivo	Desocupado	Formal	Informal	Total ocupado
Sexo					
Hombre	15,21	6,06	33,67	45,06	78,73
Mujer	34,92	7,47	28,97	28,64	57,61
Jefe de hogar					
Jefe	12,09%	4,60	37,55	45,76%	83,31
No jefe	33,74%	8,10	27,43	30,72%	58,15
Unido					
Unido	21,19	5,18	32,67	40,96%	73,62
No unido	30,21	8,39	29,69	31,72%	61,40
Nivel educativo					
Sin título	32,32	4,76	14,25	48,67%	62,92
Bachiller	28,44	7,70	28,69	35,17	63,86
Técnico/tecnólogo	13,19	9,70	49,81	27,29	77,10
Universitario	8,45	9,60	75,2	6,71	81,95
Posgrado	4,80	4,42	86,96	3,82	90,78
Región					
Atlántico	27,55	6,42	28,64	37,39	66,04
Oriental	24,84	6,73	29,45	38,99	68,43
Central	25,41	7,06	32,51	35,02	67,53
Pacífico	26,13	7,38	30,85	35,64	66,49
Bogotá	18,58	6,25	46,94	28,23	75,17
Edad					
Edad1	73,56	5,15	4,60	16,69	21,29
Edad2	36,02	11,71	24,58	27,68	52,26
Edad3	17,52	9,93	40,95	31,60	72,55
Edad4	13,78	7,45	42,87	35,91	78,77
Edad5	20,10	4,68	31,87	43,35	75,22

Fuente: elaboración propia a partir de datos de la GEIH 2017 del DANE.

comprendido por personas entre los 35 años y la edad de pensión, es decir, los adultos⁹.

La tabla 1 presenta los datos descriptivos de las variables del modelo. En términos generales, se observa que los hombres, los jefes de hogar y los unidos tienen mayor porcentaje de ocupados dentro de cada grupo. Asimismo, a mayor nivel educativo, mayor proporción de ocupados y en la formalidad. La región con mayor proporción de ocupados es Bogotá, pero las demás presentan un porcentaje similar. Por último, según grupo de edad, los más jóvenes tienen el mayor porcentaje de inactivos y la

menor proporción de ocupados. En los demás grupos aumenta la proporción de ocupados, aunque en la informalidad; esto sucede excepto en el grupo de edad 5, que tiene más personas formales que informales (véase el anexo 2, matriz de correlaciones).

Modelo econométrico

El análisis incluye dos tipos de modelos de elección discreta: un probit y un logit multinomial. Los modelos de elección discreta permiten el estudio de fenómenos que no presentan un comportamiento continuo sino de elección específica entre dos alternativas (probit y logit) o más (logit multinomial).

9 El programa utilizado para los modelos econométricos fue Stata versión 2014.

El primer modelo por estimar mide la probabilidad de estar ocupado frente a estar desempleado; por ello la variable dependiente toma valores de 1 y 0, ocupado y desempleado, respectivamente. Así, el modelo probit de empleo y empleo formal será¹⁰:

$$Pr(\text{empleado}) = f(\text{sexo}, \text{título}, \text{jefe de hogar}, \text{estado civil}, \text{dpto.}, \text{edad})$$

Luego, el modelo probit para determinar la probabilidad de estar en un empleo formal será:

$$Pr(\text{formal}) = f(\text{sexo}, \text{título}, \text{jefe de hogar}, \text{estado civil}, \text{dpto.}, \text{edad})$$

Por último, se estima el modelo logit multinomial:

$$ML(\text{Inactivo}, \text{desempleado}, \text{formal}, \text{informal}) = f(\text{sexo}, \text{título}, \text{jefe de hogar}, \text{estado civil}, \text{dpto.}, \text{edad})$$

RESULTADOS

Los resultados se presentan en tres partes: a) los resultados del modelo de empleo, b) el modelo de formalidad laboral y c) el logit multinomial entre desempleo, formalidad e informalidad. La tabla 2 presenta los resultados de la estimación del modelo de empleo. En primer lugar, se observa que la variable *sexo* es estadísticamente significativa en el modelo general y en los grupos de edad. Ser hombre aumenta la probabilidad de estar ocupado respecto a ser mujer. Sin embargo, esta probabilidad disminuye con la edad; los coeficientes más altos están en los más jóvenes y disminuyen conforme aumentan los grupos de edad.

En segundo lugar, ser jefe de hogar aumenta la probabilidad de estar ocupado; sin embargo, su coeficiente disminuye con la edad desde el grupo *edad2* al *edad5*. En tercer lugar, la variable *unido*, que representa a quienes están casados o en unión libre, muestra una relación positiva y estadísticamente significativa con la probabilidad de estar ocupado, excepto en el grupo *edad1*. Al igual que

en la anterior variable, los coeficientes disminuyen con la edad, es decir, los más jóvenes unidos tienen más probabilidad de estar empleados, y esta es más baja conforme aumenta la edad, cuando pueden adquirir más peso en las variables los solteros, divorciados o viudos.

En cuarto lugar, en el modelo general la variable *educación*, que refiere al último título obtenido, arrojó coeficientes con signo negativo y estadísticamente significativos en todos los niveles educativos, excepto posgrado, que tuvo efecto positivo. Por grupo de edad, se observa que en *edad1* para los bachilleres y técnicos/tecnólogos disminuye la probabilidad de estar ocupado respecto a personas sin ningún título. En este grupo de edad no se analizan los universitarios y posgrado porque es muy baja la proporción de personas con estos títulos (0,01% universitario y 0% posgrado). En el grupo *edad2*, la variable *educación* no tiene efecto estadísticamente significativo en ningún nivel. A partir del grupo *edad3*, el nivel educativo de posgrado aumenta la probabilidad de estar ocupado. En el grupo *edad5*, el nivel educativo universitario tiene efecto negativo y estadísticamente significativo en la probabilidad de estar ocupado. En este grupo de edad, el nivel educativo de posgrado está positiva y significativamente relacionado con la probabilidad de estar empleado.

Por el lado de la *experiencia*, se observa que esta es estadísticamente significativa en el modelo general; la experiencia potencial con signo positivo muestra que los primeros años de experiencia aumentan la probabilidad de estar ocupado, mientras que la experiencia potencial al cuadrado con signo negativo permite inferir que, a medida que aumenta dicha variable, disminuye su efecto (rendimientos decrecientes). Esto mismo ocurre en los grupos de *edad1* y *edad5*. Por su parte, en el caso de los grupos 2 y 3 solo es significativa la experiencia potencial, pues los primeros años de experiencia aumentan la probabilidad de estar ocupado, y la acumulación en el tiempo (experiencia potencial al cuadrado) no afectaría esta relación. En el grupo de *edad4* esta variable no obtuvo significancia estadística.

¹⁰ El anexo 1 contiene los detalles formales de los modelos por estimar.

Respecto de la variable *región*, que analiza a cada región frente a la parte Atlántica, en el modelo general se observa que en todas disminuye la probabilidad de estar ocupados, excepto de Bogotá (sin significancia estadística). Por grupos de edad, en la región Oriental tiene un impacto negativo en los grupos 1 y 5, mientras en los demás no tienen significancia estadística. En la región Central la probabilidad de estar ocupados solo aumenta en el grupo de edad 3, mientras en el grupo 2 y 4 no es estadísticamente significativo y en los demás disminuye. La región Pacífica disminuye esta probabilidad en los grupos 1, 4 y 5, y no lo aumenta en ninguno. En el caso de Bogotá aumenta la probabilidad de estar ocupado en todos los grupos, menos en el 5, y no es estadísticamente significativa en los grupos 4 y 2.

Por último, la relación entre la edad y la probabilidad de estar ocupado es positiva en el modelo general. Se observa cómo por grupos de edad¹¹ el valor de los coeficientes aumenta cuando se incrementa la edad. Este resultado es consistente con lo planteado en el apartado descriptivo de la metodología: el desempleo de los jóvenes es superior al de las demás edades (figura 2), y a mayor edad hay mayor probabilidad de estar ocupado (excepto grupo dos que no tienen significancia estadística).

El segundo modelo estimado es un probit que toma en cuenta a los ocupados y evalúa la probabilidad de la formalidad laboral (tabla 3)¹². La variable *sexo* es estadísticamente significativa en todos los modelos. El signo es negativo en todos, lo cual indica que la probabilidad de estar ocupado en un empleo formal aumenta si se es mujer, aunque este efecto es mayor en los adultos. En la variable *jefe de hogar* se observa que ser el jefe aumenta la probabilidad de estar en un empleo formal en todos los grupos, menos en edad 1, cuyo coeficiente no es estadísticamente significativo. Por su parte, el efecto de la variable *unido* —estar casado o en unión

libre— tiene un efecto negativo y estadísticamente significativo en la probabilidad de tener un empleo formal en todos los modelos.

Por su parte, la variable *educación* evidencia todo su efecto en este modelo probit de empleo formal: el nivel educativo aumenta la probabilidad de estar en un empleo formal en el modelo general y en todos los grupos de edad. Esta probabilidad se incrementa con el nivel educativo, como lo refleja el mayor valor de los coeficientes cuando aumenta la educación. En el caso de la *experiencia*, esta muestra en el modelo general una relación positiva en los primeros años y negativa cuando se acumula a través del tiempo (efectos no lineales), lo que implica que a medida que aumenta la experiencia, disminuye su efecto en el empleo formal. En el caso de los grupos de edad se observa lo contrario en los grupos 3 y 4, donde el efecto positivo sobre el empleo formal llega cuando más se acumula experiencia, lo que indica que los primeros años de experiencia pueden ayudar a que se ocupen las personas, pero para que sea en el sector formal se requiere de cierta cantidad acumulada.

En cuanto a la variable *región*, todas las regiones aumentan la probabilidad de estar en un empleo formal respecto a la región Atlántico. En general, los coeficientes disminuyen a medida que aumenta la edad, por lo que la influencia de las regiones va disminuyendo en la empleabilidad formal. La región Central y Bogotá presentan los mejores coeficientes. Respecto a la variable *edad*, en el modelo general se observa que todas son estadísticamente significativas y que el coeficiente solo es negativo en los más jóvenes (grupo 1). A medida que aumenta la edad, el coeficiente aumenta en este grupo, pero desciende en el grupo 3 y 4.

Leídos en conjunto, los resultados de los dos modelos arrojan varias pistas. En primer término, la educación no aumenta la probabilidad de estar ocupado. Su efecto se presenta en aumentar la probabilidad de estar ocupado en un empleo formal. El ser hombre incrementa la probabilidad de estar ocupado, pero no en la formalidad, aunque esto se da en mayor medida en los más adultos. Las mujeres tienen mayor probabilidad de estar en

11 El grupo 5 se omite por colinealidad.

12 Viáfara y Uribe (2009), Quiñones (2010) y Uribe, Ortiz y García (2007) han incluido en esta especificación la variable *experiencia potencial*. En el presente artículo fue también agregada y los resultados obtenidos son en esencia similares. Dichos resultados no se adjuntan en este texto, pero están disponibles por demanda.

Tabla 2.

Modelo de empleo

Variables	General		Edad1		Edad2		Edad3		Edad4		Edad5	
	dy/dx	P>z	dy/dx	P>z	dy/dx	P>z	dy/dx	P>z	dy/dx	P>z	dy/dx	P>z
Sexo	0,0310946	0,000	0,0656711	0,000	0,06488	0,000	0,0508353	0,000	0,0425417	0,000	0,0152098	0,000
Jefe de hogar	0,0299939	0,000	0,0489714	0,005	0,0730213	0,000	0,0563194	0,000	0,0402043	0,000	0,0151682	0,000
Unido	0,0164367	0,000	-0,0403726	0,010	0,0359644	0,000	0,0276717	0,000	0,0223702	0,000	0,0089116	0,000
Educación												
Bachiller	-0,0056551	0,000	-0,0890839	0,000	-0,0086191	0,071	0,0010832	0,782	-0,0005018	0,898	-0,0025868	0,038
Técnico/tecnólogo	-0,0046708	0,003	-0,1659826	0,000	0,0003457	0,958	0,0028748	0,560	0,0023286	0,652	-0,0058521	0,001
universitario	-0,0055757	0,002			0,0066717	0,515	0,0028147	0,633	-0,0011236	0,864	-0,0065809	0,000
Posgrado	0,0295905	0,000			0,060661	0,072	0,0493219	0,000	0,0223289	0,003	0,0216449	0,000
Expe_potencial	0,002765	0,000	-0,0285076	0,000	0,0054173	0,000	0,0028808	0,020	-0,0006982	0,723	0,0016703	0,000
Expe_potencial2	-0,0000309	0,000	0,0032348	0,000	-0,0000385	0,694	-0,0000168	0,753	0,0000663	0,242	-0,0000162	0,000
Región												
Oriental	-0,0058032	0,000	-0,022398	0,014	0,0013028	0,787	0,0056951	0,105	-0,0002137	0,949	-0,0105989	0,000
Central	-0,0070006	0,000	-0,0487216	0,000	-0,0007905	0,853	0,0104106	0,000	-0,0021651	0,455	-0,0118771	0,000
Pacífica	-0,0097267	0,000	-0,0176993	0,055	-0,0046742	0,348	-0,0017228	0,626	-0,0095981	0,005	-0,0123627	0,000
Bogotá	-0,0010342	0,600	-0,0773291	0,000	0,0085182	0,259	0,0195976	0,000	0,00198	0,716	-0,0061242	0,006
Edad												
Edad1	-0,0040327	0,100										
Edad2	0,0114698	0,000										
Edad3	0,0159867	0,000										
Ddad4	0,0175878	0,000										
	Number of obs = 389,858		Number of obs = 12,023		Number of obs = 46,903		Number of obs = 66,199		Number of obs = 51,297		Number of obs = 213,435	
	Wald chi2(17)= 10249,76		Wald chi2(11)= 319,033		Wald chi2(13)= 994,14		Wald chi2(13)= 1405,50		Wald chi2(13)= 972,76		Wald chi2(13)= 1465,45	
	Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000	
	Pseudo R2 = 0,0553		Pseudo R2 = 0,0320		Pseudo R2 = 0,0243		Pseudo R2 = 0,0339		Pseudo R2 = 0,0369		Pseudo R2 = 0,0188	
	Log pseudolikelihood = -97133,086		Log pseudolikelihood = -4908,9312		Log pseudolikelihood = -19038,379		Log pseudolikelihood = -20562,847		logpseudolikelihood = -12763,855		Log pseudolikelihood = -39544,353	

dy/dx= efecto marginal

Fuente: elaboración propia a partir de datos de la GEIH 2016 del DANE.

Tabla 3.

Modelo de empleo formal

Variable	General		Edad1		Edad2		Edad3		Edad4		Edad5	
	dy/dx	P>z										
Sexo	-0,0272855	0,000	-0,0911687	0,000	-0,0503395	0,000	-0,0118617	0,002	-0,0160438	0,000	-0,0261313	0,000
Jefe de hogar	0,0210074	0,000	0,0344441	0,069	0,0540958	0,000	0,0184887	0,000	0,0198146	0,000	0,0206778	0,000
Unido	-0,0155296	0,000	-0,0305542	0,164	-0,0480122	0,000	-0,0221999	0,000	-0,0088642	0,033	-0,0085322	0,000
Educación												
Bachiller	,1923509	0,000	0,1926008	0,000	0,2748258	0,000	0,2730832	0,000	0,2394932	0,000	0,1454291	0,000
Técnico/tecnólogo	0,3631328	0,000	0,3638521	0,000	0,47112316	0,000	0,4392207	0,000	0,4129039	0,000	0,3036475	0,000
Universitario	0,6570506	0,000			0,7049823	0,000	0,6725557	0,000	0,6614868	0,000	0,6387718	0,000
Posgrado	0,7027754	0,000					0,7104832	0,000	0,6899269	0,000	0,6930853	0,000
Expe_potencial	0,0044341	0,000	0,0098735	0,052	0,0076352	0,001	-0,0122144	0,000	-0,0124323	0,001	-0,0004344	0,492
Expe_potencial2	-0,000121	0,000	0,0007633	0,116	0,0005749	0,000	0,0006718	0,000	0,0003701	0,000	-0,0000619	0,000
Región												
Oriental	0,0220379	0,000	0,0371357	0,001	0,0626693	0,000	0,0377472	0,000	0,0110256	0,068	0,0077645	0,005
Central	0,0616384	0,000	0,0379313	0,000	0,0949061	0,000	0,0887509	0,000	0,0568698	0,000	0,0482237	0,000
Pacífica	0,0460352	0,000	0,0568055	0,000	0,073606	0,000	0,0534885	0,000	0,0329538	0,000	0,0361749	0,000
Bogotá	0,1725087	0,000	0,2845299	0,000	0,2811236	0,000	0,20145	0,000	0,1431032	0,000	0,1449068	0,000
Edad												
Edad1	0,1206704	0,000										
Edad2	0,121151	0,000										
Edad3	0,0992138	0,000										
Edad4	0,0671657	0,000										
	Number of obs = 361,029		Number of obs = 10,226		Number of obs = 39,996		Number of obs = 59,686		Number of obs = 47,612		Number of obs = 203,458	
	Wald chi2(17)= 72,141,64		Wald chi2(11)= 1094,55		Wald chi2(12)= 4770,97		Wald chi2(13)= 9905,86		Wald chi2(13)= 9196,53		Wald chi2(13)= 42769,31	
	Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000	
	Pseudo R2 = 0,1938		Pseudo R2 = 0,1031		Pseudo R2 = 0,1016		Pseudo R2 = 0,1597		Pseudo R2 = 0,1833		Pseudo R2 = 0,2139	
	Log pseudolikelihood = -201052,91		Log pseudolikelihood = -5104,0113		Log pseudolikelihood = -24882,787		Log pseudolikelihood = -34356,319		Log pseudolikelihood = -26780,615		Log pseudolikelihood = -109082,18	

dy/dx= Efecto marginal

Fuente: elaboración propia a partir de datos de la GEIH 2016 del DANE.

un empleo formal, especialmente en los grupos de adultos. El ser jefe de hogar aumenta la probabilidad de estar ocupado y en un empleo formal. El estado civil *unidos* (casado o unión libre) aumenta la probabilidad de estar ocupado, pero no de tener un empleo formal. La experiencia potencial indica que los primeros años de experiencia ayudan sobre todo a ocuparse en el sector formal. En cuanto a la regiones, Pacífico es la más afectada, pues presenta resultados negativos en la probabilidad de estar ocupado, y aunque tiene resultados positivos en la probabilidad de estar ocupado en sector formal, sus coeficientes son menores en comparación con las demás regiones.

La tabla 4 presenta los resultados de estimación de un modelo logit multinomial para analizar la influencia de la educación en la probabilidad de estar desempleado, ocupado en el sector formal y ocupado en el sector informal. En primer lugar, la variable *sexo* es estadísticamente significativa en todas las opciones de la variable dependiente. En el caso de desempleados y formales presenta signo negativo, lo que indica que el ser mujer aumenta la probabilidad de estar desempleado o en la informalidad (es equivalente a afirmar que ser hombre disminuya esta probabilidad). Por su parte, ser hombre acrecienta la probabilidad de estar ocupado, pero lo hace en mayor medida en el sector informal.

La variable *jefe de hogar* disminuye la probabilidad de estar desempleado y en el sector informal, y aumenta la probabilidad de estar ocupado en el sector formal. Este resultado es consistente con lo obtenido en los modelos anteriores y confirma la mayor presión que tienen los jefes de hogar para entrar en el mercado de trabajo y ocuparse, y lo hacen en mayor medida en la formalidad laboral. En cuanto al estado civil, estar casado o en unión libre aumenta la probabilidad de estar en la informalidad, y disminuye la probabilidad de estar desocupado y en el sector formal.

La variable *educación* disminuye la probabilidad de estar ocupado en el sector informal, Asimismo, la educación aumenta la probabilidad

de estar desempleado y de estar en el sector formal. Un resultado consistente es que el valor de los coeficientes del nivel educativo aumenta en las cuatro opciones de la variable dependiente, es decir, al mejorar una escala de nivel educativo, aumenta la probabilidad de estar desempleado —excepto en posgrado— y de estar en el sector formal. Al tiempo, disminuye en mayor medida la probabilidad de estar informal. El salario de reserva puede ser la explicación para este resultado: los más educados poseen mayor salario de reserva, por lo que optan por permanecer en desempleo hasta encontrar un trabajo formal; por el contrario, los menos educados con menor salario de reserva deben ocuparse más rápido, y con ello aumenta la probabilidad de aceptar un empleo informal. Otro factor es la segmentación en el mercado de trabajo que explica la alta informalidad y la tendencia de los menos educados y los jóvenes a caer en dicho estado.

En el caso de la *experiencia*, esta aumenta la probabilidad de ocuparse en el sector formal en los primeros años e informal a medida que acumulan más experiencia. Lo anterior tiene un efecto positivo en el sector formal en experiencia y negativo en experiencia al cuadrado, y caso contrario en el informal, pues los primeros años de experiencia tienen impacto positivo en la formalidad, pero a medida que acumulan más experiencia, tienen más probabilidad de estar en la informalidad.

El análisis de la variable *región* muestra que en Bogotá hay mayor probabilidad de estar ocupado en el sector formal, y menor en el Pacífico. Vivir en la región Pacífico aumenta la probabilidad de estar desempleado; por el contrario, vivir en Bogotá la disminuye. Finalmente, en cuanto a la variable *edad*, se observa que para los más jóvenes (*edad1* y *edad2*) aumenta la probabilidad de estar desempleado y en la informalidad (*grupo1*), y disminuye la probabilidad de estar ocupado tanto en el sector formal. En los grupos 2, 3 y 4 disminuye la probabilidad de estar informal, y se incrementa la de estar desempleado y formal. Con la edad aumenta la probabilidad de estar ocupado en el sector formal, en un efecto similar al de la educación.

Tabla 4.

Variable	Desempleo		Formal		Informal	
	dy/dx	P>z	dy/dx	P>z	dy/dx	P>z
Sexo	-0,0315638	0,000	-0,0102087	0,000	0,0417725	0,000
Jefe de hogar	-0,0320687	0,000	0,0372994	0,000	-0,0052307	0,001
Unido	-0,0173594	0,000	-0,0050358	0,001	0,0223953	0,000
Educación						
Bachiller	0,0039931	0,001	0,1765876	0,000	-0,1805807	0,000
Técnico/tecnólogo	0,0034061	0,035	0,3351115	0,000	-0,3385177	0,000
Universitario	0,0049469	0,009	0,6075377	0,000	-0,6124846	0,000
Posgrado	-0,0310493	0,000	0,6834912	0,000	-0,6524419	0,000
Expe_potencial	-0,0031379	0,000	0,0059408	0,000	-0,0028029	0,000
Expe_potencial2	0,0000392	0,000	-0,0001383	0,000	0,0000991	0,000
Región						
Oriental	0,0048426	0,000	0,0216288	0,000	-0,0264714	0,000
Central	0,0059588	0,000	0,058708	0,000	-0,0646668	0,000
Pacífica	0,0090443	0,000	0,0402691	0,000	-0,0493133	0,000
Bogotá	0,0010062	0,612	0,1666872	0,000	-0,1676934	0,000
Edad						
Edad1	0,0186074	0,000	-0,0710983	0,000	0,0524908	0,000
Edad2	0,0162968	0,000	0,0436205	0,000	-0,0599173	0,000
Edad3	0,0061469	0,004	0,0480025	0,000	-0,0541494	0,000
Edad4	0,0028245	0,105	0,0288297	0,000	-0,0316543	0,000
Number of obs = 389,858						
Wald chi2(34) = 70906,64						
Prob > chi2 = 0,0000						
Pseudo R2 = 0,1539						
Log pseudolikelihood = -298010,19						

Fuente: elaboración propia a partir de datos de la GEIH 2016 del DANE.

CONCLUSIONES

El desempleo juvenil es uno de los grandes problemas de la economía global. El caso colombiano representa un interesante caso de estudio: por un parte, el país tiene la mayor tasa de desempleo de jóvenes con alta educación dentro de los países latinoamericanos; por otra, la tasa de desempleo de jóvenes con baja educación es de las más bajas de la región. El panorama de las mujeres es el más complejo: las mujeres jóvenes con más educación son quienes presentan más problemas en su empleabilidad; por su parte, la informalidad laboral impacta más a los jóvenes y a las mujeres.

Este artículo analizó la relación entre la educación y el empleo formal. Se estimaron dos tipos de modelo de elección discreta: un probit para

analizar la probabilidad de estar ocupado y otro para determinar la probabilidad de estar en un empleo formal. También se estimó un logit multinomial para analizar conjuntamente la elección de estar en inactividad, desempleo, formalidad e informalidad. Los resultados muestran que la educación no es un determinante de la probabilidad de estar ocupado (modelo 1) y que su efecto se manifiesta en la probabilidad de tener un empleo formal (modelo 2). En efecto, al evaluar la informalidad, se observa que la educación determina en gran medida la probabilidad de estar en un empleo formal y que esta aumenta con el nivel educativo y la edad. Así, los individuos con mayor educación presentan mayor probabilidad de estar en el sector formal, al igual que los adultos.

El modelo logit multinomial confirma este hallazgo: en primer lugar, en la relación educación-desempleo se encuentra que el valor de los coeficientes de los niveles educativos técnico y universitario es mayor que el de los otros niveles de educación, hecho que puede estar asociado a la existencia de un mayor salario de reserva; en segundo lugar, la relación entre educación y empleo formal es positiva y estadísticamente significativa.

La alta informalidad laboral en Colombia puede ser la principal razón para que los más educados tengan problemas en emplearse. La informalidad responde a diversos factores, por ejemplo, a las vacantes precarias que generan las empresas de baja productividad, o a la disposición de los más educados de esperar por un mejor empleo. Este último puede ser el caso de los jóvenes con mayor salario de reserva. Los resultados en mención confirman anteriores, hallazgos como los de García (2010), Galvis (2012), Guataquí, García y Rodríguez (2010) y Sánchez (2015) en cuanto a la relación negativa entre nivel educativo e informalidad, y también los de Flores (2002), en cuya investigación encuentra que los jóvenes comienzan su ciclo de vida laboral en el sector informal y se mueven hacia el formal cuando acumulan experiencia.

La literatura consultada evidencia que los canales de búsqueda de empleo pueden explicar este efecto de la educación, porque el uso de canales informales ayuda a ocupar más personas (véase el apartado “Estudios para el caso colombiano”). Sin embargo, el uso de dichos canales genera problemas en el emparejamiento del mercado de trabajo, porque estos canales son familiares, de amigos, etc., que corresponden a círculos socioeconómicos. Así, personas que acceden a mayores niveles educativos —que con mayor probabilidad provienen de familias con mayores ingresos— cuentan con redes que facilitan el acceso a información de empleos formales. En contraste, las personas que no cuentan con estas redes, por sus condiciones socioeconómicas desfavorables, solo acceden a información de empleos informales. Este problema

de asimetría de información se podría superar si en ambos casos se usaran canales formales de búsqueda de empleo. De esta forma, los jóvenes con educación superior, pero sin redes que faciliten la información de empleos de mayor calidad, podrían ocuparse en empleos formales en mayor medida.

Los resultados por sexo confirman que las mujeres enfrentan mayores problemas en el mercado de trabajo, tienen mayor probabilidad de estar desempleadas, aunque menor probabilidad de estar ocupadas en la informalidad. No obstante, la brecha de ocupación es aún muy alta. Por otro lado, la experiencia constituye una variable relevante para estar ocupado y estarlo en el sector formal, principalmente los primeros años de experiencia laboral (por su tendencia decreciente), lo que puede ser pensado como un aspecto de política pública.

En síntesis, el análisis de la relación entre educación, edad y empleo reporta que la primera no supone mayores probabilidades de estar empleado en los más jóvenes. El efecto de la educación aparece en la calidad del empleo. En efecto, la relación entre educación y formalidad laboral es positiva. La educación aumenta la probabilidad de tener empleo formal en todas las edades. Una posible explicación a las altas tasas de desempleo juvenil de los más educados es la alta informalidad del mercado de trabajo de Colombia.

Por último, los resultados de este estudio plantean un reto para las políticas públicas de empleo; en específico, llama la atención sobre los efectos de la educación en la calidad del empleo y en el empleo de los jóvenes. Así, políticas como el salario mínimo y las políticas activas de empleo (PAMT), que a priori tienen un efecto positivo en el empleo, deben orientarse a combatir la informalidad, para poder absorber la creciente oferta de jóvenes educados. Una futura línea de investigación que se abre a partir de este artículo puede analizar los efectos de la educación en el primer empleo, y las políticas de generación de habilidades blandas en trabajadores formales e informales y su efecto en el mercado de trabajo.

REFERENCIAS

1. Attanasio, O., Kugler, A. y Meghir, C. (2007). *Effects of youth training in developing countries: Evidence from a Randomized Training Program in Colombia*. Londres: University College London.
2. Baert, S., Cockx, B. y Verhaest, D. (2013). Overeducation at the start of the career: Stepping stone or trap? *Labour Economics*, 25, 123-140. <https://doi.org/10.1016/j.labeco.2013.04.013>
3. Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to schooling*. Nueva York: National Bureau of Economic Research.
4. Belan, P. y Chéron, A. (2014). Turbulence, training and unemployment. *Labour Economics*, 27, 16-29. <https://doi.org/10.1016/j.labeco.2014.01.001>
5. Brunner, B., & Kuhn, A. (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics*, 27(3), 705-738. <https://doi.org/10.1007/s00148-013-0494-4>
6. Camargo, A. L. y Mora, H. G. (2011). La economía informal en Villavicencio. *Revista Finanzas y Política Económica*, 3(2), 83-104.
7. Cárdenas, M. y Mejía, C. (2007). *Informalidad en Colombia: nueva evidencia*. Bogotá: Fedesarrollo.
8. Cárdenas, R. J., Guataquí, R. J. y Montaña, D. J. (2015). Metodología para el análisis de demanda laboral mediante datos de internet: el caso colombiano. *Revista de Economía del Rosario*, 18(1), 93-126. DOI: [dx.doi.org/10.12804/rev.econ.rosario.18.01.2015.03](https://doi.org/10.12804/rev.econ.rosario.18.01.2015.03)
9. Carroll, D. y Tani, M. (2013). Over-education of recent higher education graduates: New Australian panel evidence. *Economics of Education Review*, 32, 207-218. <https://doi.org/10.1016/j.econedurev.2012.10.002>
10. Castillo Robayo, C. D., Da Silva Bichara, J. y Pérez-Trujillo, M. (2017). Retornos salariales para Colombia: un análisis cuantílico. *Apuntes del CENES*, 36(63), 211-246. <https://doi.org/10.19053/01203053.v36.n63.2017.5830>
11. Cheron, A., Hairault, J. O. y Langot, F. (2013). Life-cycle equilibrium unemployment. *Journal of Labor Economics*, 31(4), 843-882. <https://doi.org/10.1086/669941>
12. Cockx, B. y Ghirelli, C. (2016). Scars of recessions in a rigid labor market. *Labour Economics*, 41, 162-176. <https://doi.org/10.1016/j.labeco.2016.05.009>
13. Correa Ramírez, J. D. y Montoya Arbeláez, J. A. (2013). El valor del capital humano: una aproximación desde el enfoque del ingreso para Colombia, 2001-2009. *Cuadernos de Economía*, 32(60), 467-504.
14. Del Río Carrasquilla, F., Yáñez Contreras, M. y Pérez Arroyo, J. (2012). Duración del desempleo y eficiencia de la búsqueda de empleo en Cartagena, Colombia. *Cuadernos de Economía*, 31(58), 145-171.
15. Departamento Administrativo Nacional de Estadística (DANE) (2009a). Metodología informalidad. Gran Encuesta Integrada de Hogares. Recuperado de http://www.dane.gov.co/files/investigaciones/boletines/ech/ech_informalidad/metodologia_informalidad.pdf
16. Departamento Administrativo Nacional de Estadística (DANE) (2009b). Medición del empleo informal. Trimestre móvil septiembre-noviembre de 2009. Recuperado de http://www.dane.gov.co/files/investigaciones/boletines/ech/ech_informalidad/bolet_ech_informalidad_sep_nov2009.pdf
17. Dolado, J. J., Jansen, M. y Jimeno, J. F. (2009). On-the-Job Search in a Matching Model with Heterogeneous Jobs and Workers. *The Economic Journal*, 119(534), 200-228. <https://doi.org/10.1111/j.1468-0297.2008.02210.x>

18. Dolton, P y Vignoles, A. (2000). The incidence and effects of overeducation in the UK graduate labour market. *Economics of Education Review*, 19(2), 179-198. [https://doi.org/10.1016/S0272-7757\(97\)00036-8](https://doi.org/10.1016/S0272-7757(97)00036-8)
19. Flores, C. E. (2002). *El sector informal urbano en Colombia 1984-2000*. Bogotá: Universidad de los Andes.
20. Galvis, L. A. (2012). *Informalidad laboral en las áreas urbanas de Colombia*. Bogotá: Banco de la República.
21. García, G. A. (2010). *Evolución de la informalidad laboral en Colombia: determinantes macro y efectos locales*. Bogotá: Departamento Nacional de Planeación.
22. García, J. I., Quevedo, C. H. y Fonnegra, J. B. (2004). *Determinantes de las decisiones en el mercado laboral: La decisión de ser informal en Colombia 1988-2000*. Cali: Universidad del Valle-CIDSE.
23. García, J. I., Quevedo, C. H. y Fonnegra, J. B. (2004). *Determinantes de las decisiones en el mercado laboral: La decisión de ser informal en Colombia 1988-2000*. Cali: Universidad del Valle -CIDSE.
24. Gaure, S., Røed, K. y Westlie, L. (2012). Job search incentives and job match quality. *Labour Economics*, 19(3), 438-450. <https://doi.org/10.1016/j.labeco.2012.04.001>
25. Giuliano, L. (2013). Minimum wage effects on employment, substitution, and the teenage labor supply: Evidence from personnel data. *Journal of Labor Economics*, 31(1), 155-194. <https://doi.org/10.1086/666921>
26. Gómez Meneses, F. E. y Zárate Camelo, M. A. (2011). Gasto público en educación frente al comportamiento de los principales agregados económicos en Latinoamérica. *Revista Finanzas y Política Económica*, 3(1), 27-37.
27. González, C. G., Mora, J. J. y Cuadros, A. F. (2014). Oportunidades educativas y características familiares en Colombia: un análisis por cohortes. *Revista de Economía del Rosario*, 17(1), 157-187. DOI: [dx.doi.org/10.12804/rev.econ.rosario.17.01.2014.06](https://doi.org/10.12804/rev.econ.rosario.17.01.2014.06)
28. González-Quintero, N. I. y Báez, N. A. D. (2015). Determinantes y perfiles de la participación laboral en Colombia en el periodo 2002-2013. *Revista de Economía del Rosario*, 18(1), 5-59. <https://doi.org/10.12804/rev.econ.rosario.18.01.2015.01>
29. Guataquí, J. C., Cárdenas, J. y Montaña, J. (2014). La problemática del análisis laboral de demanda en Colombia. *Perfil de Coyuntura Económica*, 24. <https://doi.org/10.17533/udea.pece.n24a04>
30. Guataquí, J. C., García, A. F. y Rodríguez, M. (2010). El perfil de la informalidad laboral en Colombia. *Perfil de Coyuntura Económica*, 16.
31. Lamo, A., Messina, J. y Wasmer, E. (2011). Are specific skills an obstacle to labor market adjustment? *Labour Economics*, 18(2), 240-256. <https://doi.org/10.1016/j.labeco.2010.09.006>
32. Lasso, F. (2011). La dinámica del desempleo urbano en Colombia. *Borradores de Economía*, 667.
33. Léné, A. (2011). Occupational downgrading and bumping down: The combined effects of education and experience. *Labour Economics*, 18(2), 257-269. doi: <https://doi.org/10.1016/j.labeco.2010.11.007>
34. López, H. (1996). *Ensayos sobre economía laboral colombiana*. Medellín: Fonade y Carlos Valencia Editores.
35. Martínez, H. F. (2003). *¿Cuánto duran los colombianos en el desempleo y el empleo?: un análisis de supervivencia*. Bogotá: Universidad de los Andes.
36. McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, 20(3), 387-418.
37. McGuinness, S. y Wooden, M. (2009). Overskilling, job insecurity, and career mobility. *Industrial Relations: a Journal of Economy and Society*, 48(2), 265-286. <https://doi.org/10.1111/j.1468-232X.2009.00557.x>

38. Melo, L. A. y Ballesteros, C. A. (2013). Creación, destrucción y reasignación del empleo en el sector manufacturero colombiano. *Revista de Economía Institucional*, 15(18), 281-311.
39. Meroni, E. C. y Vera-Toscano, E. (2017). The persistence of overeducation among recent graduates. *Labour Economics*, 48, 120-143. <https://doi.org/10.1016/j.labeco.2017.07.002>
40. Mondragón-Vélez, C., Peña, X. y Wills, D. (2010). Labor market rigidities and informality in Colombia. *Economía*, 11(1), 65-95. <https://doi.org/10.1353/eco.2010.0009>
41. Montt, G. (2017). Field-of-study mismatch and overqualification: Labour market correlates and their wage penalty. *IZA Journal of Labor Economics*, 6(1). <https://doi.org/10.1186/s40172-016-0052-x>
42. Mora, J. J. (2008). Sobre-educación en el mercado laboral colombiano. *REI*, 19, 293-309.
43. Mora, J. J. y Ulloa, M. P. (2011). Calidad del empleo en las principales ciudades colombianas y endogeneidad de la educación. *Revista de Economía Institucional*, 13(25).
44. Mroz, T. A., y Savage, T. H. (2006). The long-term effects of youth unemployment. *Journal of Human Resources*, 41(2), 259-293. <https://doi.org/10.3368/jhr.XLI.2.259>
45. Núñez Méndez, J. A. (2002). *Empleo informal y evasión fiscal en Colombia*. Bogotá: Departamento Nacional de Planeación.
46. Organización Internacional del Trabajo (OIT) (2013). *Tendencias mundiales del empleo juvenil 2013: una generación en peligro*. Ginebra: Autor.
47. Organización Internacional del Trabajo (OIT) (2015). *Tendencias mundiales del empleo juvenil*. Ginebra: Autor.
48. Organización para la Cooperación y el Desarrollo Económicos (OCDE) (2008). *OECD Factbook 2008: Economic, Environmental and Social Statistics*. París: Autor.
49. Pellizzari, M. (2010). Do friends and relatives really help in getting a good job? *ILR Review*, 63(3), 494-510. <https://doi.org/10.1177/001979391006300307>
50. Prada, C. F. (2012). Análisis de los flujos de trabajadores y la segmentación laboral en Colombia. *Revista Ensayos Sobre Política Económica (ESPE)*, 30(68). <https://doi.org/10.32468/Espe.6802>
51. Quiñones, D. (2010). Canales de búsqueda de empleo y duración del desempleo en Colombia. *Perfil de Coyuntura Económica*, 16, 133-154.
52. Riddell, W. C. y Song, X. (2011). The impact of education on unemployment incidence and re-employment success: Evidence from the US labour market. *Labour Economics*, 18(4), 453-463. <https://doi.org/10.1016/j.labeco.2011.01.003>
53. Rivas, G. (2004). Dinámica del empleo industrial en Colombia 1977-1999. *Cuadernos PNUD-Ministerio de la Protección Social*, 4.
54. Sánchez Torres, R. M. (2015). Identificación y caracterización de los trabajadores pobres en Colombia, 2002-2012. *Revista de Economía Institucional*, 17(32). <https://doi.org/10.18601/01245996.v17n32.10>
55. Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, 51(1), 1-17.
56. Uribe, J. I. y Ortiz, C. H. (2006). *Informalidad laboral en Colombia 1988-2000: evolución, teorías y modelos*. Cali: Universidad del Valle.

57. Uribe, J., Ortiz, C. y Correa, J. (2004). *Determinantes de las decisiones en el mercado laboral: la decisión de ser informal en Colombia*. Cali: Universidad del Valle.
58. Uribe, J. I., Ortiz, C. H. y García, G. A. (2007). La segmentación del mercado laboral colombiano en la década de los noventa. *Revista de Economía Institucional*, 9(16).
59. Viáfara, C. A. y Uribe, G. (2009). Duración del desempleo y canales de búsqueda de empleo en Colombia. *Revista de Economía Institucional*, 11(21), 139-160.
60. Zenou, Y. (2015). A dynamic model of weak and strong ties in the labor market. *Journal of Labor Economics*, 33(4), 891-932.

ANEXOS

Anexo 1. Detalles del modelo probit y logit multinomial

Este modelo se ha usado para determinar la participación en el mercado de trabajo, segmentación laboral y determinar la probabilidad de estar empleado (Espino y Sauval, 2016; Paz, 2013; Sánchez, 2015; González y Báez, 2015; González, Mora y Cuadros, 2014; Uribe y Ortiz, 2006). De este modo, los individuos pueden estar en el empleo (modelo 1) y en la formalidad (modelo 2).

$$\text{Empleado} = \begin{cases} 0 & \text{si empleado} \leq 0 \\ 1 & \text{si empleado} > 0 \end{cases}$$

En este modelo el interés está en estimar $E\left[\frac{\text{empleado}}{X}\right]$ en términos del vector de los parámetros, donde el valor esperado es la probabilidad de $E\left[\frac{\text{empleado}}{X}\right] = P(\text{empleado} = \frac{1}{X})$. Donde:

$$\text{Empleado}^* = \beta'x + \varepsilon_i$$

$$P[\text{empleado}_i = 1] = P[\beta'x_i + \varepsilon_i > 0] = F(\beta'x_i) \quad [1]$$

Donde ε_i es el termino de error o perturbación aleatoria del modelo; $F(\beta'x_i)$ es la distribución normal acumulada de ε_i condicionado a X_i ; es decir, si $F(\beta'x_i)$ es una función de distribución acumulada de ε_i , este modo la probabilidad estará dada por:

$$\begin{aligned} P[\text{empleado}_i = 1] &= P[\text{Empleado}^* = \beta'x_i = \beta'X + \varepsilon_i > 0] \\ &= P[\varepsilon_i > -\beta'X] = 1 - F[-\beta'X] \\ &= F[\beta'X] \end{aligned} \quad [2]$$

Por tanto, el modelo probit estaría determinado como:

$$\int_{-\infty}^{\beta'X} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \theta(\beta'X) \quad [3]$$

Donde θ representa la función de distribución normal acumulada. Al momento de analizar los resultados, hay que resaltar que los parámetros del modelo no son los efectos marginales de este, que normalmente se analizan en el modelo de mínimos cuadrados ordinarios (MCO); por tanto,

para analizar los resultados, hay que determinar los efectos marginales que vendrían determinados por:

$$\frac{\partial P(\text{empleado} = 1)}{\partial X_i} = \frac{\partial [F(\beta'X)]}{\partial X_i} = f(\beta'X)\beta_i \quad [4]$$

Donde $f(\cdot)$ es la función de densidad de la distribución estándar normal asociada a la función de distribución $F(\cdot)$; por tanto, si se utiliza una distribución normal, se tendría:

$$\frac{\partial P(\text{empleado} = 1)}{\partial X_i} = \phi(\beta'X)\beta_i \quad [5]$$

Ya que se asume X_i es una variable binaria (valores entre 0 y 1) y se necesita observar el cambio de $P(\text{empleado}_i = 1|X_i)$ cuando X_i toma valores entre 0 y 1, el efecto marginal de modelo sería:

$$\begin{aligned} \partial P(\text{empleado}_i = 1|X_i) \partial X_i &= P(Y_i = 1|X_1 = 1, X_2, \dots, X_k) - P(\text{empleado}_i = 1|X_1 = 0, X_2, \dots, X_k) \\ &= F(\beta_0 + \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k) - F(\beta_0 + \beta_2 X_2 + \dots + \beta_k X_k) \end{aligned} \quad [6]$$

Donde $F(\cdot)$ es la distribución normal acumulada del modelo probit. Cabe anotar que si una variable X_i es dicotómica, como se usará regularmente en las regresiones, el análisis del efecto de la variación se mide a través de la diferencia entre los valores determinados por:

$$E\left[\frac{\text{empleado} = 1}{X_i = 1}\right] - E\left[\frac{\text{empleado} = 1}{X_i = 0}\right] \quad [7]$$

Los modelos por estimar mediante la regresión Probit son:

$$\text{empleado} = f(\text{sexo}, \text{título}, \text{jefe de hogar}, \text{estado civil}, \text{dpto.}, \text{edad})$$

$$\text{formal} = f(\text{sexo}, \text{título}, \text{jefe de hogar}, \text{estado civil}, \text{dpto.}, \text{edad})$$

Logit multinomial: el modelo logit multinomial también ha sido usado para analizar el mercado laboral (Mora y Ulloa, 2015; Uribe, Ortiz y Correa, 2006; Posso, 2010). A diferencia de los modelos de elección binaria como modelos probit y logit (elección 0 y 1), los modelos de elección múltiple analizan la elección entre varias alternativas (elección 1 hasta J) determinada por las variables explicativas. En este modelo se asigna 1 a inactivos,

2 a desempleados, 3 a ocupados formales y 4 a ocupados informales. De esta forma se evalúa cómo un individuo toma una de las cuatro alternativas de forma racional, es decir, con base en la teoría de la utilidad marginal. Es así como $U_{ij}(J-1)$ recoge esta utilidad en las diferentes alternativas y X representa las variables explicativas sobre las cuales se toma la decisión; por ello, cada individuo elige el que mayor utilidad tenga por encima de las otras opciones. Se tiene entonces:

$$U_{ij} = \beta'X_{ij} + \varepsilon_i$$

De este modo, lo que se evaluará es la probabilidad de que el individuo i tome una alternativa j para $j = 0, 1, 2 \dots J$:

$$P_{ij} = P(ML = j|X_i) \quad [8]$$

Para la elección, en este caso se tendría que:

$$ML_i \begin{cases} 1 \text{ si } U_{i1} > U_{ij} \text{ para toda } j \neq 1 \\ 2 \text{ si } U_{i2} > U_{ij} \text{ para toda } j \neq 2 \\ 3 \text{ si } U_{i3} > U_{ij} \text{ para toda } j \neq 3 \\ 4 \text{ si } U_{i4} > U_{ij} \text{ para toda } j \neq 4 \\ j-1 \text{ si } U_{i(j-1)} > U_{ij} \text{ para toda } j \neq (j-1) \end{cases}$$

El orden de los valores y de la base para la estimación es arbitraria y su interpretación dependerá de los efectos marginales, al igual que el modelo probit, además de que la interpretación depende de la variable latente ML_{ij}^* . Esta última puede interpretarse como la utilidad no observada que el individuo obtiene al elegir la alternativa j :

$$ML_{ij}^* = X_i' \beta_j + \varepsilon_{ij} \quad [9]$$

En este caso se evalúa qué probabilidad tiene el individuo de estar en uno de estos cuatro grupos dependiendo de las variables explicativas (las mismas del modelo probit). Por otra parte, en el modelo logit multinomial se supone que los términos de error $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J$ son independientes y siguen una distribución de valores extremos. Se puede demostrar que, desde esta suposición, las probabilidades tienen la forma de la distribución logística:

$$P_{ij} = P(ML_i = j|X_i) = \frac{\exp(X_i' \beta_j)}{1 + \sum_{h=1}^j \exp(X_i' \beta_h)} \quad j = 1, 2, \dots, j \quad [10]$$

Donde la probabilidad de la primera alternativa ($j = 0$) está completamente determinada por la probabilidad del resto de ellos (ya que deben sumar a la unidad). Se usa la normalización $\exp(X_i' \beta_0) = 1$; como ya se señaló, los coeficientes estimados no informan sobre el signo del efecto marginal, es decir, un coeficiente positivo de un cierto regresor no implica un aumento en la probabilidad cuando este regresor aumenta. Por ello, se debe calcular el efecto marginal con derivados o diferencias (esto depende de si es variable continua o discreta); en este caso se obtiene el efecto de un cambio de un regresor con *odds ratios*, o relaciones de riesgo relativo (la relación entre la probabilidad de dos alternativas). La razón impar de la alternativa j con la alternativa 0 y con el resto de las alternativas se puede escribir de la siguiente manera:

$$\frac{P(ML_i = j|X_i)}{P(ML_i = h|X_i)} = \frac{\exp(X_i' \beta_j)}{1} = \exp(X_i' \beta_j)$$

$$\frac{P(ML_i = j|X_i)}{P(ML_i = h|X_i)} = \frac{\exp(X_i' \beta_j)}{\exp(X_i' \beta_h)} = \exp(X_i' \beta_j - \beta_h)$$

$$h = 1, \dots, j \quad [11]$$

Los log *odds ratios* están dados por una función lineal:

$$\log(P(ML_i = j|X_i) / P(ML_i = 0|X_i)) = X_i' \beta_j$$

$$\log(P(ML_i = j|X_i) / P(ML_i = 0|h_i)) = X_i'(\beta_j - \beta_h)$$

$$h = 1, \dots, J \quad [12]$$

A partir de esto se pueden calcular los efectos marginales; sin embargo, cabe decir que este modelo presenta un inconveniente en el comportamiento: la independencia de alternativa irrelevante, que se refiere a la relación entre la probabilidad de dos alternativas diferentes que no depende del resto de las alternativas (Wooldridge, 2006). De esta forma, el modelo por estimar mediante la regresión logit multinomial es:

$$ML \text{ (Inactivo, desempleado, formal, informal)} \\ = f(\text{sexo, título, jefe de hogar, estado civil,} \\ \text{dpto., rae, edad})$$

Anexo 2. Matriz de correlación polychoric

Variable	Ocupado	Sexo	Educación	Jefe de hogar	Estado civil	Región
Ocupado	1					
Sexo	0,16397911	1				
Educación	-0,06233353	-0,15891124	1			
Jefe de hogar	0,27623756	0,33864722	-0,09325544	1		
Estado civil	0,20640892	0,13794403	-0,05619462	0,27138857	1	
Región	-0,01047031	-0,02329935	0,00922651	0,03760709	-0,08449381	1

Anexo 3. test de Wald - modelo probit de empleo

General		Edad1		Edad2		Edad3		Edad4		Edad5	
-1	[oci]sexo = 0	-1	[oci]sexo = 0	-1	[oci]sexo = 0	-1	[oci]sexo = 0	-1	[oci]sexo = 0	-1	[oci]sexo = 0
-2	[oci]jefehogar = 0	-2	[oci]jefehogar = 0	-2	[oci]jefehogar = 0	-2	[oci]jefehogar = 0	-2	[oci]jefehogar = 0	-2	[oci]jefehogar = 0
-3	[oci]estadocivil = 0	-3	[oci]estadocivil = 0	-3	[oci]estadocivil = 0	-3	[oci]estadocivil = 0	-3	[oci]estadocivil = 0	-3	[oci]estadocivil = 0
-4	[oci]1b.titulo = 0	-4	[oci]1b.titulo = 0	-4	[oci]1b.titulo = 0	-4	[oci]1b.titulo = 0	-4	[oci]1b.titulo = 0	-4	[oci]1b.titulo = 0
-5	[oci]2.titulo = 0	-5	[oci]2.titulo = 0	-5	[oci]2.titulo = 0	-5	[oci]2.titulo = 0	-5	[oci]2.titulo = 0	-5	[oci]2.titulo = 0
-6	[oci]3.titulo = 0	-6	[oci]3.titulo = 0	-6	[oci]3.titulo = 0	-6	[oci]3.titulo = 0	-6	[oci]3.titulo = 0	-6	[oci]3.titulo = 0
-7	[oci]4.titulo = 0	-7	[oci]4.titulo = 0	-7	[oci]4.titulo = 0	-7	[oci]4.titulo = 0	-7	[oci]4.titulo = 0	-7	[oci]4.titulo = 0
-8	[oci]5.titulo = 0	-8	[oci]1b.region = 0	-8	[oci]5.titulo = 0	-8	[oci]5.titulo = 0	-8	[oci]5.titulo = 0	-8	[oci]5.titulo = 0
-9	[oci]1b.region = 0	-9	[oci]2.region = 0	-9	[oci]1b.region = 0	-9	[oci]1b.region = 0	-9	[oci]1b.region = 0	-9	[oci]1b.region = 0
-10	[oci]2.region = 0	-10	[oci]3.region = 0	-10	[oci]2.region = 0						
-11	[oci]3.region = 0	-11	[oci]4.region = 0	-11	[oci]3.region = 0						
-12	[oci]4.region = 0	-12	[oci]5.region = 0	-12	[oci]4.region = 0						
-13	[oci]5.region = 0		Constraint 4 dropped	-13	[oci]5.region = 0						
	Constraint 4 dropped		Constraint 8 dropped		Constraint 4 dropped						
	Constraint 9 dropped				Constraint 9 dropped						
	chi2(11) = 5721,54		chi2(10) = 92,43		chi2(11) = 663,81		chi2(11) = 1326,59		chi2(11) = 719,25		chi2(11) = 898,07
	Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000

Anexo 4. Test de Wald - modelo probit de formalidad

General		Edad1		Edad2		Edad3		Edad4		Edad5	
-1	[formal]sexo = 0	-1	[formal]sexo = 0	-1	[formal]sexo = 0	-1	[formal]sexo = 0	-1	[formal]sexo = 0	-1	[formal]sexo = 0
-2	[formal]jefehogar = 0	-2	[formal]jefehogar = 0	-2	[formal]jefehogar = 0	-2	[formal]jefehogar = 0	-2	[formal]jefehogar = 0	-2	[formal]jefehogar = 0
-3	[formal]estadocivil = 0	-3	[formal]estadocivil = 0	-3	[formal]estadocivil = 0	-3	[formal]estadocivil = 0	-3	[formal]estadocivil = 0	-3	[formal]estadocivil = 0
-4	[formal]1b. titulo = 0	-4	[formal]1b. titulo = 0	-4	[formal]1b. titulo = 0	-4	[formal]1b. titulo = 0	-4	[formal]1b. titulo = 0	-4	[formal]1b. titulo = 0
-5	[formal]2. titulo = 0	-5	[formal]2. titulo = 0	-5	[formal]2. titulo = 0	-5	[formal]2. titulo = 0	-5	[formal]2. titulo = 0	-5	[formal]2. titulo = 0
-6	[formal]3. titulo = 0	-6	[formal]3. titulo = 0	-6	[formal]3. titulo = 0	-6	[formal]3. titulo = 0	-6	[formal]3. titulo = 0	-6	[formal]3. titulo = 0
-7	[formal]4. titulo = 0	-7	[formal]4. titulo = 0	-7	[formal]4. titulo = 0	-7	[formal]4. titulo = 0	-7	[formal]4. titulo = 0	-7	[formal]4. titulo = 0
-8	[formal]5. titulo = 0	-8	[formal]1b. region = 0	-8	[formal]5. titulo = 0	-8	[formal]5. titulo = 0	-8	[formal]5. titulo = 0	-8	[formal]5. titulo = 0
-9	[formal]1b. region = 0	-9	[formal]2. region = 0	-9	[formal]1b. region = 0	-9	[formal]1b. region = 0	-9	[formal]1b. region = 0	-9	[formal]1b. region = 0
-10	[formal]2. region = 0	-10	[formal]3. region = 0	-10	[formal]2. region = 0						
-11	[formal]3. region = 0	-11	[formal]4. region = 0	-11	[formal]3. region = 0						
-12	[formal]4. region = 0	-12	[formal]5. region = 0	-12	[formal]4. region = 0						
-13	[formal]5. region = 0		Constraint 4 dropped	-13	[formal]5. region = 0						
	Constraint 4 dropped		Constraint 8 dropped		Constraint 4 dropped						
	Constraint 9 dropped				Constraint 9 dropped						
	chi2(11) = 31973,85		chi2(10) = 263,85		chi2(11) = 2002,51		chi2(11) = 5142,75		chi2(11) = 4696,01		chi2(11) = 19809,91
	Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000		Prob > chi2 = 0,0000

Anexo 5. Relaciones de riesgo relativo (RRR), modelo logit multinomial

Multilogit	RRR	P>z	RRR	P>z	RRR	P>z
Base outcome						
1	2		3		4	
Sexo	0,5317314	0,000	0,3360068	0,000	0,334905	0,000
Jefe de hogar	1.086.059	0,000	2.005.564	0,000	1.789.578	0,000
Estado civil	0,6254324	0,000	0,8378575	0,000	0,8952978	0,000
Título						
2	1.822.239	0,000	2.381.498	0,000	1.215.726	0,000
3	4.502.362	0,000	7.673.303	0,000	1.758.848	0,000
4	6.721.237	0,000	152.221	0,000	0,5354758	0,000
5	6.173.701	0,000	2.827.866	0,000	0,4688344	0,000
Región						
2	114.265	0,000	1.258.751	0,000	1.194.665	0,000
3	121.023	0,000	1.398.713	0,000	107.834	0,000
4	1.151.476	0,000	1.123.938	0,000	0,969668	0,056
5	1.465.611	0,000	2.635.834	0,000	1.317.691	0,000
Edad						
Edad1	0,2528076	0,000	0,0608287	0,000	0,0852291	0,000
Edad2	0,9436694	0,012	0,4230462	0,000	0,3060378	0,000
Edad3	1.845.601	0,000	1.283.854	0,000	0,8215296	0,000
Edad4	2.023.213	0,000	1.755.129	0,000	1.261.272	0,000
_cons	0,4338058	0,000	2.455.644	0,000	8.428.092	0,000
Number of obs = 311,65						
Wald chi2(45) = 84391,41						
Prob > chi2 = 0,0000						
Pseudo R2 = 0,1570						
Log pseudolikelihood = -332324,62						