



Análisis del Desempleo en Bogotá: Determinantes para 2017 Tercer Trimestre¹

Daniel Navarro Azuero
Luisa Juliana Ojeda Barrera

Resumen

Teniendo en cuenta el bajo desempeño de la economía colombiana para años recientes y la incertidumbre que rodea las perspectivas sobre su crecimiento, se ha venido generando una mayor preocupación acerca de la capacidad del país para generar y mantener el empleo de sus habitantes. Este documento centra tal preocupación en Bogotá, la ciudad más poblada del país, para donde se busca estimar qué aspectos contribuyen en mayor medida a generar desempleo, por medio de un modelo no lineal de probabilidad.

Palabras clave: Desempleo, Búsqueda de empleo, Salario de Reserva, Modelo Logit

¹ Trabajo de grado para optar por el título de Economista en la Universidad de la Sabana.
Director: Martha Misas Arango

CONTENIDO

INTRODUCCIÓN	3
MOTIVACIÓN PERSONAL	4
REVISIÓN DE LITERATURA	4
MODELO ECONÓMICO	7
MODELACIÓN ECONOMETRICA	11
• CALIDAD DE PREDICCIÓN	12
• EFECTOS MARGINALES	13
DATOS	14
• DESCRIPCIÓN DE VARIABLES	14
• METODOLOGÍA	15
RESULTADOS MODELO ECONOMETRICO	16
• EFECTOS MARGINALES	19
• CALIDAD DE PREDICCIÓN	20
CONCLUSIONES	21
REFERENCIAS	21

INTRODUCCIÓN

En macroeconomía, uno de los principales temas de estudio es el crecimiento económico, que se encuentra estrechamente ligado al desempeño del mercado laboral, cuya fortaleza puede ser estudiada a través de la tasa de desempleo. Cabe aclarar que el análisis de este indicador debe estar sujeto al conocimiento de su conformación de acuerdo al país de estudio, para poder emitir juicios acorde con los contextos apropiados.

Un claro ejemplo de la importancia de esta relación se puede observar en el caso europeo, donde los países que han logrado mantener una tasa de desempleo cercana al pleno empleo (4,5%) han alcanzado un crecimiento mayor en su PIB y un mercado más estable. Resalta el caso particular de Luxemburgo y España: para 2001, el primero tuvo una tasa de desempleo de 1,9% y el segundo de 10,5% y hoy en día Luxemburgo tiene un mejor crecimiento per cápita que España. En la literatura se atribuye este fenómeno a la mayor cantidad de personas aportando dentro de las firmas y asimismo a un crecimiento mayor del PIB (Cahuc & Zylberberg, 2004).

Teniendo en cuenta el bajo desempeño de la economía mundial para años recientes y la incertidumbre que rodea las perspectivas sobre su crecimiento, se ha venido generando una mayor preocupación acerca de la capacidad de los distintos países para generar y mantener el empleo en sus territorios. Para 2017, se espera que 3,4 millones de personas se sumen a la población en situación de desempleo, que para finales del año alcanzaría los 201 millones de personas. En este sentido, la Organización Internacional del Trabajo (OIT) señala su profunda preocupación por la situación de América Latina en particular, cuya tasa de desempleo se espera que continúe en aumento; mientras en los países desarrollados se espera que disminuya.

Así mismo, a pesar de que para Colombia la tasa de desempleo se ha mantenido más o menos constante durante los últimos 9 años, la difícil situación económica del país genera preocupación sobre su posible trayectoria en el corto plazo. Resulta imperativo reducir la tasa de desempleo, pues en 2017 esta fue del 9,4% y el crecimiento del PIB fue del 1,8%. Sin embargo, para poder establecer una estrategia de reducción es necesario entender cuáles son los determinantes del desempleo, para así poder implementar las políticas públicas correctas. (DANE, 2018)

En este sentido, se propone investigar acerca del desempleo para la ciudad de Bogotá, la más poblada de Colombia, con un 16,39% de la población total y altos niveles de desigualdad.

MOTIVACIÓN PERSONAL

Como estudiantes universitarios prontos a ingresar al mercado laboral, probablemente en la capital, nos interesa hacer un análisis de la estructura del empleo, con especial atención a las causas del desempleo. Consideramos de suma importancia entender el contexto al cual vamos a ingresar, así como la probabilidad de conseguir empleo que tiene un individuo con nuestras características en dicho mercado, para mejorar el proceso de toma de decisiones que puedan afectar nuestras oportunidades laborales.

REVISIÓN DE LITERATURA

En primer lugar, Tenjo y Ribero (1997) han realizado diversas investigaciones sobre el mercado laboral en Colombia, en las que plantean que el estudio del desempleo ha cobrado mayor importancia desde 1985 hasta la actualidad, haciendo énfasis en el aumento de la participación de la mujer en el mercado laboral, fundamentado en el mayor nivel educativo que ellas han obtenido a través de los años.

También establecen la conexión que tiene el desempleo con los ciclos económicos.

La apertura económica abrió las posibilidades de una mayor movilidad de fuerza laboral entre firmas, pero a su vez aumentó la duración de búsqueda de un trabajo por el ajuste estructural que se debe dar.

Desde otra perspectiva, se describen las fuerzas de la ley de Okun en el contexto colombiano, que implican una relación negativa entre la tasa de desempleo y la actividad económica, teniendo en cuenta que la tasa de desempleo sería cíclica y está vinculada a la evolución económica y productividad. Además, analizan tasas que no sean cíclicas y tiendan a un modelo friccional de desempleo.

En un estudio de modelos de participación laboral se encuentra que hay 5 posibles variables relevantes:

1. Medidas de ingreso no laboral, que se construyen por riqueza propia o familiar y se ha encontrado que tienen un efecto negativo en la participación.
2. Salario potencial en el mercado, que normalmente está marcado por la educación, que debería tener un efecto positivo sobre la probabilidad de ser contratado.
3. Edad, pues por el ciclo de vida de las personas, las decisiones de participar en el mercado laboral o no cambian a través del tiempo. Por ejemplo, los jóvenes le dan mayor importancia al estudio en cierto periodo de tiempo y hacen un mayor énfasis en buscar trabajo en otro periodo.

4. Número de hijos pequeños, que normalmente se expresa por medio de una variable dummy donde por lo general 1 hace referencia a que hay existencia de hijos pequeños en el núcleo familiar. Esta variable tiene un efecto negativo en la participación de las madres.
5. En el contexto colombiano se puede incluir el servicio doméstico, que puede ser un sustituto de la madre, dándole a esta la posibilidad de participar más activamente en el ámbito laboral.

Adicionalmente, se divide el desempleo en tres categorías para crear la tasa de desempleo: el desempleo cíclico, el desempleo friccional (que se genera por la demora de ajuste entre la oferta y la demanda del trabajo) y el desempleo estructural (la no coincidencia entre la oferta y demanda laboral). El desempleo cíclico evidencia la relación que hay entre desempleo y crecimiento del PIB, por ejemplo en Colombia en 1994, que tuvo una tasa de crecimiento de 5,5%, con una tasa natural de desempleo de 4,9%, pero en 1997 tuvo un retroceso del crecimiento del PIB del 1,22% con un desempleo de 13,6%.

Finalmente, los autores postulan características inexplicables del desempleo en países latinoamericanos:

1. El desempleo se centra en zonas urbanas.
2. La mayor parte del desempleo es juvenil.
3. El desempleo típico está bien educado.
4. La mayoría de los desempleados son solteros.
5. La mayoría de los desempleados dependen económicamente de sus familias.

(Tenjo & Ribero, 1997)

Desde otro punto de vista, un análisis interesante es el trabajo que se hizo en la ciudad de Cali, donde se buscaron los determinantes de la duración del desempleo en el periodo de 1988 a 1998 y se llegó a formar un modelo con coeficientes significativos, así que fue posible hacer inferencia, por medio del uso de econometría, basado en el búsqueda de empleo. Se encontró que los determinantes para la duración del desempleo en Cali eran: la posición en el hogar, el sexo, los ingresos no laborales del trabajador, el nivel educativo, la experiencia y el coeficiente de variación salarial. Esto abre la posibilidad de hacer una investigación sobre los determinantes del desempleo, pero en basados en la ciudad de Bogotá (Castellar & Uribe, 2006).

Por otro lado, Bernal & Núñez (1998) hacen una distinción entre el desempleo coyuntural, que es aquel producto del contexto económico preponderante en el periodo de análisis; y el desempleo natural, que se refiere al generado por desbalances habituales entre oferta y demanda en el mercado laboral (estructural), así como al mínimo nivel de desempleo posible en una economía sana, ocasionado por el cambio regular de empleo de los individuos o la búsqueda de empleo de nuevos participantes del mercado (friccional).

Para el caso de Colombia, los aumentos de la tasa de desempleo se relacionan con mayores niveles de la tasa de cambio, crecientes costos no salariales y la brecha de oferta y demanda de la mano de obra no calificada. Adicionalmente, se resalta la ampliación de la participación de la mujer en el mercado laboral. También se pone de relieve la brecha entre la oferta y demanda de mano de obra calificada que, dada la inversión tecnológica realizada en el momento de estudio, requiere la sustitución de mano de obra no calificada por calificada.

Con respecto a la duración del desempleo, se propone una relación estrecha con las características de los individuos, el ciclo económico y los beneficios del desempleo, y se plantea el imperativo de estudiarla con base en su influencia en la probabilidad de encontrar empleo, en la generación de pobreza, en el salario de reserva y en la moral de los individuos desempleados. En este sentido, se tiene que a pesar de que la mayor tasa de desempleo cobija a los individuos jóvenes y las mujeres, también son estos quienes tienen mayor probabilidad de encontrar empleo.

Finalmente, los resultados de la modelación econométrica, medidos con diferentes variables sugieren que:

- Los individuos jóvenes tienen mayor probabilidad de salir del desempleo.
- Los hombres solteros tienen una menor probabilidad que los casados de salir del desempleo, situación que se revierte en el caso de las mujeres, por cuestiones de tiempo esperado de estadía en el empleo.
- Los individuos con mayor número de dependientes tienen un mayor tiempo de búsqueda, así como una mayor probabilidad de encontrar empleo.
- Los ingresos no laborales relajan la búsqueda del individuo y aumentan su salario de reserva.
- La probabilidad de enganche en el sector construcción es la mayor entre las ramas de actividad, mientras en los sectores de electricidad, agua, gas y financiero la probabilidad es la menor.
- Los individuos con mayor educación tienen mayor probabilidad de salir del desempleo, pero existen disparidades de género, pues para las mujeres la probabilidad ha venido aumentando, mientras que para los hombres ha venido disminuyendo. (Bernal & Núñez, 1998)

Desde otro ángulo, Viáfara y Uribe (2009) hablan acerca de la duración del desempleo y cómo esta depende de dos efectos particulares: el efecto aspiraciones, que se basa en los determinantes tanto del salario de reserva (por el lado de la oferta) como del salario potencial (por el lado de la demanda), que atienden a características de los individuos, como lo son el género, la jefatura del hogar, la educación, la experiencia, entre otros; y el efecto oportunidades, que se relaciona con el acceso a cierta información no pública, que se adquiere a través de las relaciones sociales del individuo. En consecuencia, la persona decidirá cuál de los efectos predomina en sus características y de acuerdo a esto enfocará su búsqueda de empleo hacia canales formales o informales.

En este sentido, se describe cómo los estudios en Colombia se han centrado en estudiar el efecto aspiraciones, mediante una aproximación a través de los factores causales del salario de reserva y el capital humano.

Finalmente, al realizar mediciones de duración a través del estadístico de Kaplan-Meier se encontró que cuando se controla por el canal de búsqueda (de acuerdo con el efecto predominante) los buscadores de empleo más eficientes son los “jefes de hogar, hombres, personas con menor educación, mayor experiencia”, esto en cuanto a una menor duración del desempleo, pues la mayor urgencia por emplearse y la menor calificación hace que estos individuos se definan un salario de reserva bajo y de esta manera acepten ofertas laborales de baja calidad. Otro hallazgo importante es el efecto positivo de la educación sobre la probabilidad de salir del desempleo, aunque se aclara que este puede depender de la muestra utilizada en las estimaciones (Viáfara & Uribe, 2009).

Por su parte, Tenjo, Misas, Contreras, & Gaviria (2012) estudiaron no solo la probabilidad de estar en desempleo sino su duración, donde los modelos econométricos arrojaron los siguientes resultados:

- Los hombres tienen menor duración del desempleo que las mujeres.
- Las personas casadas y los jefes del hogar tienen procesos de búsqueda de empleo más cortos.
- Teniendo en cuenta la variable edad y su forma cuadrática se tiene que la duración del desempleo disminuye hasta cierto punto, a partir del cual empieza a aumentar.
- Los años de educación, igualmente incluyendo su forma cuadrática, tienen el mismo efecto que la edad.
- El ingreso del resto de la familia, una proxy de los recursos de los que dispone el desempleado para su búsqueda, aumenta la duración de la búsqueda de empleo.

MODELO ECONÓMICO

Tradicionalmente, la teoría neoclásica de oferta de trabajo ignora los costos tanto monetarios como de tiempo que implica buscar un empleo y sostiene que no hay personas desempleadas sino que optan por el ocio en lugar de trabajar formalmente; es decir, asumen un contexto de información perfecta, donde las ofertas de trabajo y el salario único de mercado están disponibles al público, quien solo debe escoger cuánto y dónde trabajar, no buscar.

Este modelo, aunque útil para ciertos análisis de mercado laboral, se aleja de verdadera estructura del empleo en el entorno real, donde existe información imperfecta y los agentes deben considerar sus opciones para seleccionar un nuevo trabajo. Este es el escenario que considera el modelo de búsqueda de empleo, que conlleva a que el individuo se proponga mejorar su bienestar por medio del salario a recibir (Cahuc & Zylberberg, 2004).

De acuerdo con dichos autores la estrategia óptima para una persona que se encuentra en busca de trabajo es elegir un salario que represente el pago mínimo que estaría dispuesta a aceptar por la prestación de sus servicios a una firma; este es llamado el salario de reserva (w_r) y se determina con base en el entorno económico, los subsidios de desempleo y la tasa en que se reciben ofertas de trabajo, de manera que maximice los beneficios netos futuros derivados de la búsqueda de empleo. En este orden de ideas, se acepta la oferta de trabajo cuya remuneración (w) esté por encima del salario de reserva.

$w \geq w_r$ Se acepta la oferta de trabajo

$w < w_r$ Se rechaza la oferta de trabajo

Posteriormente, se hace uso del modelo simple de búsqueda de empleo propuesto por Tenjo, Misas, Contreras, & Gaviria, (2012), de donde se tomarán todas las ecuaciones y que se describe detalladamente a continuación:

En el proceso de búsqueda de empleo, se asume que todos los trabajadores son idénticos y que no conocen los salarios ofrecidos por las firmas, sino la distribución acumulada de ofertas salariales en un momento establecido, que es la misma para todos (French & Taber, 1986). La función de densidad que describe esta distribución es $g(w)$, que sigue $g: (\mu, \sigma)$.

Con esto en mente, la probabilidad de que un individuo acepte la oferta de trabajo se define como:

$$Prob(w \geq w_r) = 1 - G(w_r)$$

Y su complemento, la probabilidad de que un individuo rechace la oferta de trabajo:

$$Prob(w < w_r) = G(w_r) = \int_0^{w_r} g(w) dw$$

Dado que los individuos no conocen las cualidades de las vacantes ofrecidas, estas pueden verse como un proceso aleatorio (Mortensen, 1970). En este sentido, el valor esperado del número de ofertas a recibir antes de aceptar alguna viene dado por:

$$D = E(L) = \frac{1}{1 - G(w_r)}$$

En particular, D sería la duración esperada de la búsqueda de empleo, sí y solo sí se asume que se recibe una oferta laboral por periodo, de lo contrario, la duración no sería D sino una función de esta (Tenjo et al., 2012). Asumiendo que es posible observar el tiempo que toma encontrar un trabajo y la distribución del salario que este trabajo ofrece, la única fuente de heterogeneidad del modelo sería el momento de las ofertas y la forma de la distribución de los salarios, problema que se supera al asumir la oferta por periodo (French & Taber, 1986).

Si $1 - G(w_r)$ es la probabilidad de aceptar la oferta de trabajo, su inversa será el lapso de tiempo en que las ofertas fueron rechazadas por no cumplir con la condición de salario.

Como no es observable, el salario esperado (w_e) que recibiría el individuo al obtener una oferta de trabajo aceptable sería:

$$w_e = E(w | w \geq w_r) = \frac{\int_{w_r}^{\infty} wg(w)dw}{1 - G(w_r)}$$

La definición del salario de reserva puede aproximarse como el valor esperado descontado del máximo retorno que puede obtenerse al recibir una nueva oferta de trabajo, es decir, el ingreso futuro por concepto de salario en la oferta aceptada (Y) menos los costos incurridos en la búsqueda por parte del individuo (C), quién determina si es rentable o no invertir en el mercado laboral (McCall, 1970). Debe tenerse en mente que los costos comprenden tanto directos como de oportunidad.

Los costos directos pueden verse como todos aquellos relacionados con el proceso de búsqueda, como lo son la impresión y distribución de hojas de vida, circular por la ciudad o inscribirse en plataformas de búsqueda de empleo, entre otros. Estos costos no tienden a variar y pueden considerarse fijos para cada periodo (F) (Tenjo et al., 2012). También un aumento en el coste de búsqueda disminuye el salario de reserva del buscador, ya que el individuo deberá implementar una parte del salario de reserva mayor para poder seguir en la búsqueda de trabajo, entre más bajo sean estos costos el salario de reserva ser más elevado y el individuo tenderá a aceptar una oferta de trabajo en un tiempo mayor (García & Rivera, 2007).

Por otro lado, los costos de oportunidad se refieren a los ingresos que no recibe el individuo por periodo dado que está buscando empleo en lugar de aceptar cualquier oferta por debajo de su salario de reserva y trabajar. Estos costos no son fijos, sino que se incrementan a medida que se alarga la duración de la búsqueda ($h(D)$) (Tenjo et al., 2012).

En este sentido y manteniendo el supuesto de que se recibe una oferta de trabajo por periodo, se pueden expresar los costos como función de la duración de la búsqueda de empleo:

$$C = FD + h(D) = C(w_r) \quad \text{y} \quad \frac{\partial C}{\partial w_r} > 0$$

Teniendo en mente esta relación, puede inferirse que a medida que los costos de la búsqueda aumentan, su duración disminuye, ya que dado un momento del tiempo, al enfrentar tan altos costos, al individuo le resulta más rentable abandonar la búsqueda de empleo y entrar a hacer parte los “empleados desmotivados”, personas que ya no buscan empleo activamente, pues puede enfrentar “restricciones de liquidez que hagan que solo pueda financiar el coste de búsqueda durante un intervalo limitado de tiempo” (García-Perez, 1997). En contraste, si se enfrentan bajos costos de búsqueda, el individuo seguirá buscando hasta encontrar una oferta cuyo salario exceda su salario de reserva; el tiempo hasta que esto suceda es lo que se conoce como desempleo friccional (McCall, 1970).

Los beneficios de la búsqueda se refieren a los ingresos que percibe el trabajador, medibles como función de w_e :

$$Y = Y(w_r) \quad \text{con} \quad \frac{\partial Y}{\partial w_r} > 0 \quad \& \quad \frac{\partial^2 Y}{\partial^2 w_r} < 0$$

La persona enfrenta un problema económico al fijar su salario de reserva, pues mientras más alto sea, más tiempo tendrá que pasar antes de encontrar una oferta aceptable, pero también mayor será su salario esperado una vez empleado (Mortensen, 1970).

Así, el individuo fija su salario de reserva maximizando el retorno descrito en páginas anteriores de la siguiente manera:

$$V(B) = \int_D^N Y(w_r)e^{-rt} dt - \int_0^D C(w_r)e^{-rt} dt$$

$$V(B) = \frac{1}{r} [Y(w_r)(e^{-rD} - e^{-rN}) - C(w_r)(1 - e^{-rD})]$$

Donde N , siguiendo la suposición de que se recibe una oferta por periodo, es el horizonte para la toma de decisiones, es decir, que la búsqueda del individuo termina y este inicia su vinculación al mercado laboral luego rechazar N ofertas que no cumplían la condición de ofrecer un salario mayor al de reserva (McCall, 1970).

En la literatura se encuentra que el punto o salario óptimo, que en este caso sería (w_r^*) , se genera igualando el beneficio marginal del individuo por esperar otra oferta de trabajo más adelante y con el costo marginal de la búsqueda relacionado a generar otra oferta de trabajo adicional y, junto con la duración esperada, es función de la distribución de ofertas laborales, la tasa de descuento, el horizonte para la toma de decisiones y un vector de otras variables (X) (McCall, 1970).

$$\frac{\partial V(B)}{\partial w_r} = 0 \quad \rightarrow \quad w_r^* = k(g, r, N, X)$$

$$D = d(g, r, N, X)$$

Estas otras variables, además de características propias del individuo, pueden contener factores relevantes como las condiciones del mercado laboral, a las que el desempleado debe adaptarse en su proceso de búsqueda (McCall, 1970).

MODELACIÓN EONOMÉTRICA

Para los objetivos de esta investigación, se hace necesario el uso de un modelo de respuesta cualitativa o modelo de probabilidad, donde la variable respuesta es binaria y solo puede adquirir los valores de 1 y 0, dependiendo de si el evento estudiado ocurre o no; en este caso particular el evento se refiere a estar desempleado. El propósito detrás de este tipo de modelos es encontrar la probabilidad de que tal evento suceda. (Gujarati & Porter, 2010) & (Scott, 1997)

Por lo tanto, la variable y sigue la distribución discreta de probabilidad de Bernoulli. De tal forma que la probabilidad de que el evento ocurra se define como $P_i = Prob(y_i = 1)$ y la probabilidad de que el evento no ocurra como $(1 - P_i) = Prob(y_i = 0)$. Su función de densidad es:

$$f(y_i) = P_i^{y_i}(1 - P_i)^{1-y_i} \quad (1)$$

y sus propiedades son:

$$E[y_i] = P_i \quad (2)$$

$$Var[y_i] = P_i(1 - P_i) \quad (3)$$

(Scott, 1997)

Debido principalmente al problema de acotamiento, se hace necesario el uso de un modelo de probabilidad que asuma P_i como una función no lineal de las variables explicativas y sea estimado por medio de máxima verosimilitud. Dicha función no lineal se selecciona como una función de distribución acumulada, ya que este tipo de funciones son acotadas, continuas y miden probabilidad. Así, se evalúa la función de distribución escogida en el punto $x_i'\beta$ y se obtiene:

$$P_i = F(x_i'\beta) \quad (4)$$

Las funciones más utilizadas para este propósito son la normal (modelo probit) y la logística (modelo logit)². Para efectos de esta investigación se hará uso de la función de distribución acumulada logística³.

El objetivo econométrico es maximizar la función de log-verosimilitud para estimar el vector de parámetros β y la varianza σ^2 asociados a la forma funcional definida en la ecuación (4), que en este caso es la logística.

² Es posible comparar los coeficientes estimados de los modelos logit (L) y probit (P) por medio de la siguiente fórmula: $\beta_L \approx \sqrt{Var(e_L|x)}\beta_P \approx \sqrt{\pi^2/3}\beta_P \approx 1.81\beta_P$

³ $F(t) = \frac{1}{[1+\exp(-t)]}$

La estimación del modelo se lleva a cabo a partir de un método de optimización no lineal que maximiza la función de verosimilitud. La ecuación (5) presenta la función de verosimilitud, la cual corresponde a la función de densidad conjunta.

$$L(\beta|y, x) = f(y_1, y_2, \dots, y_i|P_i) \quad \forall i = 1, \dots, n \quad (5)$$

Asumiendo independencia en el muestreo esta se puede reescribir como:

$$L(\beta|y, x) = \prod_{i=1}^n f(y_i|P_i) \quad (6)$$

Y reemplazando la función de densidad de la distribución Bernoulli ya mencionada se tiene:

$$L(\beta|y, x) = \prod_{i=1}^n P_i^{y_i} (1 - P_i)^{1-y_i} \quad (7)$$

De acuerdo con la definición $P_i = F(x_i' \beta)$ que supera el problema de acotamiento se tiene:

$$L(\beta|y, x) = \prod_{i=1}^n F(x_i' \beta)^{y_i} (1 - F(x_i' \beta))^{1-y_i} \quad (8)$$

La ecuación (10) es el resultado de la aplicación de la función logaritmo a la función de verosimilitud⁴.

$$\ln[L] = \ln \prod_{i=1}^n [F(X_i' \beta)^{y_i} (1 - F(X_i' \beta))^{1-y_i}] \quad (9)$$

$$\ln[L] = \sum_{i=1}^n y_i \ln[F(X_i' \beta)] + (1 - y_i) \ln[1 - F(X_i' \beta)] \quad (10)$$

La maximización de la ecuación (10) permite alcanzar el objetivo econométrico de estimar el vector de parámetros β y la varianza σ^2 . La naturaleza no lineal de la ecuación (10) requiere para su maximización de un procedimiento de optimización no lineal tipo Newton-Raphson, que consiste en una aproximación lineal, a partir de una expansión de Taylor de grado dos alrededor del vector $\hat{\beta}$, cuyos valores iniciales, de acuerdo con Dhrymes (1978), pueden obtenerse a través del estimador OLS del modelo lineal. (Scott, 1997)

Posterior a la estimación de los coeficientes se tiene que la probabilidad estimada es:

$$\hat{P}_i = F(x_i' \hat{\beta}) \quad (11)$$

CALIDAD DE PREDICCIÓN:

Se procede a evaluar la precisión del modelo, es decir, qué tan eficientemente clasifica las observaciones en ambas categorías posibles: $y_i = 1$ o $y_i = 0$, teniendo en cuenta las observaciones reales.

⁴ Puede aplicarse la función logaritmo gracias a su propiedad de monotonidad, que significa que no cambia el valor donde se encuentra la solución de maximización.

Debe generarse una variable indicadora \hat{y}_i :

$$\hat{y}_i = \begin{cases} 1 & \text{si } \hat{P}_i > c \\ 0 & \text{si } \hat{P}_i \leq c \end{cases} \quad 0 < c < 1 \quad (12)$$

Donde c es un umbral de clasificación que generalmente toma el valor de 50%. De manera similar, se construye la variable w_i :

$$w_i = \begin{cases} 1 & \text{si } y_i = \hat{y}_i & \text{pronóstico correcto} \\ 0 & \text{si } y_i \neq \hat{y}_i & \text{pronóstico incorrecto} \end{cases} \quad (13)$$

Para la muestra de tamaño n se define h como la proporción de predicciones acertadas, es decir, las veces que el modelo predice acertadamente en promedio:

$$\frac{1}{n} \sum_{i=1}^n w_i \quad (14)$$

(Scott, 1997)

EFECTOS MARGINALES:

Los efectos marginales miden el cambio parcial en la probabilidad del evento generado por el cambio en una variable explicativa particular, es decir, el cambio que genera un cambio en un regresor sobre la probabilidad de y_i . Así:

$$\frac{\delta \text{prob}(y_i=1|X_i)}{\delta X_{ij}} = \frac{dF(X_i'\beta)}{dX_i'\beta} \frac{\delta X_i\beta}{\delta X_{ij}} = f(X_i\beta)\beta_j \quad (15)$$

En particular, para la función logística se tiene:

$$\frac{\delta P_i}{\delta X_i} = \frac{\exp(-X_i'\beta)}{[1+\exp(-X_i'\beta)]^2} \beta_j \quad (16)$$

De acuerdo con Cameron & Trivedi (2005) existen 3 formas de calcular los efectos marginales:

- Promedio de los efectos marginales individuales: se trata de obtener el promedio de los efectos marginales de cada individuo ante el cambio de cada variable explicativa.

$$\beta_j \frac{1}{n} \sum_{i=1}^n f_i(X_i\beta) \quad (17)$$

- Efectos marginales evaluados en el promedio: consiste en calcular la media de cada una de las variables, \bar{X} , explicativas y sobre esta obtener el efecto marginal que produce el cambio en cada variable explicativa.

$$f_i(\bar{X}\beta)\beta_j \quad (18)$$

- Efectos marginales evaluados en un individuo representativo: se lleva a cabo seleccionando uno o varios individuos particulares dentro de la muestra cuyas características sean representativas y se obtienen los efectos marginales.

$$f_i(X^*\beta)\beta_j \quad (19)$$

Albarrán (2011) sostiene que los efectos marginales evaluados en el promedio no son del todo acertados para concluir sobre la muestra pues “la media no es siempre un valor representativo de una distribución” y esta alternativa excluye la relación entre los valores de las distintas variables. Con base en lo anterior, en esta investigación se calcula el promedio de los efectos marginales individuales, que según el mismo autor soluciona los inconvenientes mencionados.

DATOS

Las bases de datos utilizadas en la construcción del documento se conformaron a partir de la Gran Encuesta Integrada de Hogares (GEIH). Se tomaron las bases: características de personas, ocupados, desocupados, otros ingresos e inactivos, filtradas para la ciudad de Bogotá, para conformar la base global a usar en la estimación. Se tuvieron en cuenta los meses de Julio, Agosto y Septiembre, debido a que este trimestre del año presenta la menor cantidad de estacionalidades.

DESCRIPCIÓN DE VARIABLES:

La matriz de diseño X está conformada por:

Edad (edad): esta variable se aproxima como un indicador de la experiencia y el conocimiento del individuo acerca del mercado laboral. Adicionalmente, dado el ciclo de vida de las personas, las decisiones de las personas, no solo respecto a la participación laboral, cambian a lo largo del tiempo. Muchos de estos aspectos no se relacionan linealmente con la probabilidad de desempleo, así que debe incluirse la edad de manera cuadrática.

Educación (aeduc): esta variable se refiere al número de años de educación cursados y aprobados por cada individuo. Se busca captar el efecto de la especialización sobre la probabilidad de estar en desempleo. De la misma forma que la edad, se incluye de manera cuadrática para capturar efectos más complejos.

Reemplazo: se designan de esta manera aquellas variables que permiten explicar el salario de reserva, es decir, la urgencia de un individuo por conseguir empleo.

- Estado civil (estcivil): esta variable toma el valor de 1 si el individuo está casado o vive en unión libre y 0 de lo contrario. Un valor de 1 podría representar una mayor

carga económica para el individuo, es decir una mayor urgencia por trabajar, que se traduce en un menor salario de reserva.

- Jefe del hogar (jhogar): esta variable toma el valor de 1 si el individuo es jefe del hogar y 0 de lo contrario. Un valor de 1 podría representar una mayor carga económica para el individuo, es decir una mayor urgencia por trabajar, que se traduce en un menor salario de reserva.
- Número de menores a cargo (nmenor): esta variable se refiere al número de niños menores de 7 años que pertenecen al hogar y se entiende como una variable de responsabilidad económica. Su efecto es ambiguo pues puede ser que se tenga un mayor salario de reserva dado un mayor costo de oportunidad, o puede que se tenga un menor salario de reserva dada la urgencia por trabajar.
- Ingreso del resto de la familia (ingfam_res): esta variable se obtiene de restar el ingreso de los individuos ocupados de la familia del ingreso familiar total. En este sentido se considera como el ingreso del resto de la familia y se aproxima al salario de reserva como proxy de los recursos disponibles para financiar la búsqueda de empleo. Se toma la variable en su transformación logarítmica.
- Género (genero): esta variable toma el valor de 1 si el individuo es hombre y 0 si es mujer. Se agrega como control debido a que existen disparidades entre hombres y mujeres en cuanto al estado de desempleo.

(Tenjo et al., 2012)

METODOLOGÍA:

Inicialmente se buscó estimar el modelo por medio de un comando que replicara la modelación econométrica descrita anteriormente en un software estadístico, sin embargo, debido a un problema la estructura de los datos se optó por una alternativa diferente.

La base de datos a utilizar presenta un problema de desequilibrio de clase, es decir, el número de individuos de una clase (ocupados) es mucho mayor al de la otra (desocupados) en la variable independiente, con una relación de casi 7 a 1. Este problema genera que el algoritmo detrás de los comandos usualmente clasifique de forma equivocada las observaciones, al estar sesgado hacia la clase con mayor proporción. (Duman, Ekinici, & Tanriverdi, 2012)

De esta manera se planteó el uso del método de bootstrapping. En primer lugar se divide la muestra original en dos matrices de acuerdo con la clasificación correspondiente: ocupados (matriz grande) y desocupados (matriz pequeña). El procedimiento consiste en re muestrear de manera aleatoria sin reemplazamiento las observaciones de la matriz de ocupados para generar una sub matriz con las mismas dimensiones de la matriz de desocupados; acto seguido se pegan estas dos matrices para generar una muestra uniforme entre ocupados y desocupados y sobre esta se corre el modelo logit en el software trabajado.

Sin embargo, usar una muestra parcial puede cambiar los resultados obtenidos y hacer que no sean representativos de la muestra original. Para solucionar este problema, se repite el

procedimiento descrito anteriormente ocho mil veces y se promedian los resultados obtenidos para los estimadores, efectos marginales y calidad de predicción⁵, con el fin de generalizar los resultados a toda la muestra y poderlos analizar. La idea detrás de la repetición del muestreo se basa en el Teorema Central del Límite, según el cual, un número de variables aleatorias que tienda a infinito permite aproximar una distribución dada, en este caso, la de los parámetros. (Alvarado & Batanero, 2008)

RESULTADOS DEL MODELO ECONÓMÉRICO

La estimación econométrica consta de diez variables explicativas para la muestra total y se aplica para dos sub muestras; mujeres y hombres, por lo tanto los modelos las matrices de diseño utilizadas son:

- Muestra total:

$$X_1 = \{edad, edad^2, estcivil, jhogar, genero, aeduc, aeduc^2, nmenor, lingfam_res, lingfam_res^2\}$$

- Sub muestra mujeres:

$$X_2 = \{edad, edad^2, estcivil, jhogar, aeduc, aeduc^2, nmenor, lingfam_res, lingfam_res^2\}$$

- Sub muestra hombres:

$$X_3 = \{edad, edad^2, estcivil, jhogar, aeduc, aeduc^2, nmenor, lingfam_res, lingfam_res^2\}$$

En el análisis de los betas se tiene en cuenta que los estimadores vienen de un modelo de probabilidad no lineal, por lo tanto su signo indica la dirección del efecto de la variable explicativa del modelo. En las siguientes tablas se encuentran los coeficientes con su respectivo p-valor y desviación estándar, seguidos de los efectos marginales y la calidad de predicción.

⁵ La literatura existente hace referencia al método de bootstrapping para cálculo de distintos estadísticos y de coeficientes de regresión a través de desarrollos más complejos que los presentados en este documento; la literatura sobre el cálculo de efectos marginales a través de este método es escasa y poco precisa. Dadas las limitaciones del proyecto en cuanto a tiempo y conocimientos de econometría más avanzada se optó por promediarlos de la manera descrita en la metodología y se deja abierta la discusión acerca de su precisión.

Tabla 1

Muestra Total				
	Coefficiente	P-valor	Desviación Estándar	Efectos Marginales
Intercepto	3.0944	0.0136	0.7189	3.0944
Edad	-0.1785	0.0011	0.0318	-0.1785
Edad^2	0.0019	0.0063	0.0004	0.0019
Estcivil	-0.3030	0.1315	0.1125	-0.3030
Jhogar	-0.5831	0.0060	0.1165	-0.5831
Género	-0.1810	0.1439	0.1113	-0.2810
Aeduc	0.1575	0.2077	0.0684	0.1575
Aeduc^2	-0.0067	0.1715	0.0027	-0.0067
Nmenor	-0.0736	0.5138	0.0923	-0.0736
Lingfam_res	0.0000	0.5064	0.0000	0.0000
Lingfam_res^2	0.0021	0.1419	0.0009	0.0021
Calidad de predicción				
True	0		1	
0	266		136	
1	151		251	
Porcentaje de aciertos del modelo			64.30%	

Tabla 2

Sub Muestra: Mujeres				
	Coefficiente	P-valor	Desviación Estándar	Efectos Marginales
Intercepto	2.1604	0.2658	1.1357	2.1604
Edad	-0.1807	0.0374	0.05210	-0.1807
Edad^2	0.0017	0.1331	0.0007	0.0017
Estcivil	-0.2649	0.3367	0.1680	-0.2649
Jhogar	-0.3627	0.2472	0.1809	-0.3627
Aeduc	0.3530	0.0741	0.0991	0.3530
Aeduc^2	-0.01470	0.0521	0.0039	-0.0148
Nmenor	-0.2187	0.2773	0.1222	-0.2187
Lingfam_res	0.0000	0.3754	0.0000	0.0000
Lingfam_res^2	0.0018	0.3379	0.0015	0.0018
Calidad de predicción				
True	0		1	
0	154		67	
1	76		145	
Porcentaje de aciertos del modelo			67.65	

Tabla 3

Sub Muestra: Hombres				
	Coefficiente	P-valor	Desviación Estándar	Efectos Marginales
Intercepto	3.1598	0.2658	0.9810	3.1598
Edad	-0.1522	0.0374	0.0455	-0.1522
Edad ²	0.0018	0.1331	0.0006	0.0018
Estcivil	-0.2754	0.3367	0.1978	-0.2754
Jhogar	-0.8905	0.2472	0.1980	-0.8905
Aeduc	0.0192	0.0741	0.0961	0.0193
Aeduc ²	-0.0009	0.0521	0.0039	-0.0009
Nmenor	0.0631	0.2773	0.1450	0.0631
Lingfam_res	0.0000	0.3754	0.0000	0.0000
Lingfam_res ²	0.0015	0.3379	0.0013	0.0015
Calidad de predicción				
True	0		1	
0	128		52	
1	58		122	
Porcentaje de aciertos del modelo			69.44	

En primer lugar, se observa que el hecho de ser hombre reduce la probabilidad de estar desempleado, acorde con la teoría, ya que las mujeres tienden a encargarse de las labores del hogar y el cuidado de los niños, haciendo su salario de reserva mayor.

Se encuentra que los coeficientes de las variables de jefe del hogar y estado civil tienen signo negativo en todas las muestras, de manera consistente con la literatura, debido a que ambas variables implican responsabilidad económica sobre otros individuos y mayores gastos, razón por la cual se tiene un menor salario de reserva al buscar con más apremio tener un empleo.

En cuanto a *edad* y *edad*² para todas las muestras, se tiene que la primera tiene un signo negativo, es decir que la probabilidad de estar desempleado disminuye a medida que la persona sea mayor. Sin embargo, complementando el análisis *edad*² tiene un signo positivo, lo que indica que esta variable llega a un punto en el que deja de disminuir y empieza a aumentar la probabilidad de estar en desempleo.

Para las variables *aeduc* y *aeduc*² en todas las muestras se puede hacer un análisis similar: la primera es positiva, dando a entender que a más años de educación mayor es la probabilidad de estar en desempleo, probablemente debido a que la mano de obra menos calificada tiende a tener un salario de reserva bajo que fácilmente es igualado o superado por la oferta del mercado, haciendo que las personas se empleen con más regularidad. Respecto al cuadrado de la variable se tiene que los años de educación llegan a un punto a partir del

cual disminuyen la probabilidad de estar en desempleo, lo que puede darse debido a que un alto nivel de especialización puede hacer que el mercado ofrezca salarios muy altos que superen el salario de reserva y hagan que el individuo acepte ofertas de trabajo más a menudo, disminuyendo la probabilidad de estar en desempleo.

Respecto a *lingfam_res*, en todas las muestras, se tiene que a mayor cantidad de recursos disponibles para la búsqueda de empleo mayor probabilidad de estar en desempleo, hecho que puede explicarse dada la menor urgencia por conseguir empleo, ya que se cuenta con recursos para el sostenimiento familiar y se tiene un salario de reserva mayor, que hará que el individuo espere más tiempo por ofertas adecuadas.

Finalmente, la variable *nmenor* tiene un efecto ambiguo entre las muestras: en general, el signo negativo de la variable indica que un mayor número de menores a cargo reduce la probabilidad de estar en desempleo, debido a la mayor urgencia por obtener recursos, traducida en un bajo salario de reserva que hace que se acepten ofertas de baja calidad para asegurar el ingreso. Este resultado es consistente para la sub muestra de mujeres. Para la sub muestra de hombres esta variable tiene un signo positivo, es decir que mayor número de menores a cargo puede aumentar la probabilidad de estar en desempleado.

EFECTOS MARGINALES:

En todas las muestras la variable que más impacto tiene es el hecho de ser jefe del hogar y afecta de manera positiva la probabilidad de estar desempleado. El mayor efecto se encuentra en la sub muestra de hombres, aumentando la probabilidad en un 89.05%; el menor efecto se tiene para la sub muestra de mujeres, en la que aumenta la probabilidad en 36.27%. Por último, para la muestra total se tiene un efecto marginal de 58.31%. En la literatura se corroboran estos resultados, ya que ser responsable del sostenimiento del hogar genera urgencia por conseguir empleo, baja el salario de reserva y hace al individuo menos exigente con las ofertas.

La siguiente variable que más pesa en el modelo son los años de educación. La muestra más afectada es la de mujeres, en que el efecto es de 35.30%. Por otra parte, si se analiza la muestra total y la sub muestra de hombres, el impacto es considerablemente menor con un efecto marginal de 15.75% y 1.93%, respectivamente. En todas las muestras los años de educación aumentan la probabilidad de estar en desempleo. En este sentido, a pesar de que el efecto años de educación aplica para ambos géneros, su severidad es mayor para las mujeres que para los hombres.

Otra variable que determina la probabilidad de estar en desempleo es el género, ser hombre disminuye la probabilidad de estar en desempleo en un 28.26%, efecto que es de esperar en el contexto de un país como Colombia, donde normalmente las mujeres se limitan a las labores del hogar.

Por otro lado la variable estado civil en las tres muestras tiene un peso parecido y el mismo efecto, disminuye la probabilidad de estar en desempleo. En la muestra total tiene un efecto

marginal de -30.30%, para las mujeres es de -26.49% y para los hombres es de -27.54%. Esto quiere decir que estar casado, sin importar el género, influye en una disminución considerable en la probabilidad de estar desempleado, pues un individuo más en el hogar aumenta los gastos, de manera que se tiende a bajar la exigencia frente a las ofertas de trabajo y se prefieren estar empleado, disminuyendo el salario de reserva.

Adicionalmente, la variable menores a cargo impacta en mayor medida a la sub muestra de mujeres con un efecto marginal de -21.87%, es decir, que entre más niños a cargo tenga la mujer, menor será la probabilidad de estar desempleada. En la literatura se tiene que las mujeres normalmente reducen su salario de reserva dada la urgencia por adquirir recursos para mantener a los niños. Así mismo en los hombres tiende a disminuir la probabilidad de estar desempleado pero en 6.31%, de manera que esta variable influye aproximadamente tres veces más en las mujeres que en los hombres.

Por otra parte, el efecto de la edad tiene una magnitud similar en los tres casos, pero llega un punto en que cambia de signo. En la muestra total el efecto marginal es -17.85%, en la sub muestra de mujeres es de -18.07% y en la sub muestra de hombres es -15.22%. Esto indica que para las personas jóvenes mayor edad disminuye la probabilidad de estar desempleado, pero para las personas mayores mayor edad contribuye a estar en desempleo.

Por último, el efecto marginal del ingreso del resto de la familia contribuye a la probabilidad de estar desempleado, pero el efecto es casi nulo. En la literatura se atribuye esta variable a la financiación que se tiene para la búsqueda de empleo.

CALIDAD DE PREDICCIÓN:

Se encontró que, en cuanto a la muestra total, el modelo predice de manera correcta en 64.30% de los casos, proporción que permite hacer inferencia sobre el modelo y concluir a partir de él. Para las sub muestras se tiene incluso un mejor nivel de predicción, para las mujeres un 67.65% y para los hombres un 69.44%.

CONCLUSIONES

Los efectos de las variables estudiadas son consistentes con la literatura y permiten una mayor comprensión de los determinantes del desempleo en Bogotá. Sin embargo, hay una variable que no tiene el desempeño esperado, el ingreso del resto de la familia. Esta variable es una proxy de los recursos que el desempleado tiene disponibles para buscar empleo, así como de la urgencia por conseguirlo. La literatura sugiere que esta variable debería elevar el salario de reserva de los individuos, haciéndolos más exigentes frente a las ofertas laborales y por lo tanto aumentando la probabilidad de estar en desempleo. Los resultados de esta investigación indican que esta variable tiene el signo esperado pero su magnitud es mínima en cuanto a cambios en la probabilidad, razón por la cual se propone un desarrollo más centrado en esta variable, que de luces sobre su verdadera naturaleza.

Por otro lado, a pesar de la importancia de conocer la probabilidad de estar en desempleo y cómo esta cambia dado el efecto de diversas variables relevantes, sería interesante y bastante útil poder analizar cuánto tiempo le toma a un individuo salir del desempleo de acuerdo con sus características, así como su incidencia.

REFERENCIAS

- Albarrán, P. (2011). Tema 3: Modelos de Elección Discreta.
- Alvarado, H., & Batanero, C. (2008). Significado del teorema central del limite en textos universitarios de probabilidad y estadística. *Estudios Pedagógicos*, XXXIV(2), 7–28. <https://doi.org/10.4067/S0718-07052008000200001>
- Bernal, R., & Núñez, J. (1998). El desempleo en Colombia: tasa natural, desempleo cíclico y estructural y la duración del desempleo, (1976-1998). *Revista ESPE*, 7–74.
- Cahuc, P., & Zylberberg, A. (2004). Labor Economics. In *Labor Economics* (pp. 107–166). [https://doi.org/10.1016/S0169-7218\(11\)02402-6](https://doi.org/10.1016/S0169-7218(11)02402-6)
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*.
- Castellar, C., & Uribe, J. (2006). Determinantes de la duración del desempleo en el área metropolitana de Cali 1988-1998. *Sociedad Y Economía*, (11), 8–38.
- DANE. (2018). Indicadores Económicos. Retrieved February 28, 2018, from <http://www.dane.gov.co/index.php/67-espanol/servicios-dane/indicadores-economicos>
- Duman, E., Ekinici, Y., & Tanriverdi, A. (2012). Comparing alternative classifiers for database marketing: The case of imbalanced datasets. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2011.06.048>

- French, E., & Taber, C. (1986). The Handbook of Labour Economics. In O. Ashenfelter & D. Card (Eds.), *The Handbook of Labour Economics Vol.2* (Vol. 2, pp. 537–617).
- García-Perez, J. I. (1997). Las tasas de salida del empleo y el desempleo en España (1978-1993). *Investigaciones Económicas*, 18(July), 29–53. Retrieved from <ftp://www.funep.es/InvEcon/paperArchive/Ene1997/v21i1a2.pdf>
- García, J., & Rivera, L. (2007). Canales de búsqueda de empleo y duración del desempleo en el mercado laboral colombiano 2003. *Sociedad Y Economía*, 13, 153–173.
- Gujarati, D., & Porter, D. (2010). *Econometría*.
- McCall, J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics*, 84(1), 113–126.
- Mortensen, D. T. (1970). Job Search, the Duration of Unemployment, and the Phillips Curve. *The American Economic Review*, 60(5), 847–862.
<https://doi.org/10.2307/1818285>
- Scott, J. (1997). Regression models for categorical and limited dependent variables. *Regression Models for Categorical and Limited Dependent Variables*.
- Tenjo, J., Misas, M., Contreras, A., & Gaviria, A. (2012). Modelos de duración del desempleo en Colombia, 12, 1–55.
- Tenjo, J., & Ribero, R. (1997). Participación, desempleo y mercados laborales en Colombia. *Archivos de Macroeconomía*, 1–78.
- Viáfara, C., & Uribe, J. (2009). Duración del desempleo y canales de búsqueda de empleo en Colombia. *Revista de Economía Institucional*, 2(21), 139–160. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84879821896&partnerID=tZOtx3y1>