

Duración del desempleo en Colombia

Jaime Tenjo Galarza

Martha Misas Arango

Alfredo Contreras Eitner

Alejandro Gaviria Jaramillo

Julio de 2012

- I. Introducción
- II. Revisión de Literatura
- III. Modelo simple de búsqueda de empleo
- IV. Estimación no paramétrica de la duración del desempleo
 1. Estimador de Kaplan –Meier
 2. Resultados
- V. Modelación econométrica
 1. Modelo de duración del desempleo
 - a. Conceptos básicos
 - b. Estimación a través del modelo paramétrico AFT
 2. Modelo de probabilidad
- VI. Conjunto de información relevante en la estimación econométrica.
 - A. Descripción de variables
 - B. Construcción de las variables asociadas a las condiciones del mercado referentes a duración del desempleo y salarios por ocupación.
- VII. Resultados de la estimación de los modelos de duración y de probabilidad.
 - A. Resultados de la estimación de los modelos de duración
 - B. Resultados de los modelos de probabilidad
- VIII. Incidencia del desempleo

I. Introducción

En general la tasa de desempleo de una economía se considera el mejor indicador sintético de lo que sucede en el mercado laboral. Sin embargo, a pesar de sus ventajas, dicho indicador no arroja información sobre muchos aspectos de lo que sucede en dicho mercado. Una de estas falencias tiene que ver con el hecho de que una misma tasa de desempleo puede generarse bajo diferentes condiciones y un cambio en su nivel se puede deber a cambios en cualquiera de las condiciones que concurren a su determinación. Más específicamente, una determinada tasa de desempleo se puede deber a una situación en la cual los miembros de la fuerza de trabajo quedan desempleados con mucha frecuencia (alta incidencia), pero duran desempleados poco tiempo (baja duración); o a que quedan desempleados con baja frecuencia pero duran buscando empleo mucho tiempo. (Por ejemplo, una tasa mensual de desempleo de 12% se puede deber a que 1% de los trabajadores quedan desempleados cada mes, y duran buscando empleo 12 meses, o a que 6% de los trabajadores quedan desempleados cada mes, pero consiguen un empleo nuevo en 2 meses). Distinguir entre estos dos tipos de situaciones es importante tanto para el diseño de políticas laborales adecuadas, como para lograr un mejor entendimiento del funcionamiento del mercado laboral.

Una forma de aproximarse al entendimiento de esta dinámica es investigando tanto la probabilidad de que una persona se encuentre desempleada, como la duración del desempleo y los factores que la determinan. Estos son temas ampliamente estudiados en la literatura sobre economía laboral, pero que en el caso Colombiano no han recibido la atención que ameritan.

Este artículo se centra en la estimación de la duración del desempleo y la probabilidad de estar en desempleo para Colombia. El objetivo fundamental es proporcionar información detallada sobre las características de los procesos de búsqueda de empleo y del desempleo en Colombia, incluyendo la estimación de la *incidencia* del desempleo, un elemento relevante para el diseño de políticas de mercado laboral. La estimación de los modelos se basa en la información contenida en la Encuesta Nacional de Hogares para cabeceras municipales correspondiente al tercer trimestre (Agosto, Septiembre y Octubre) de 2010.

El artículo está dividido en ocho secciones. La primera es esta introducción. En la segunda se presenta la revisión de literatura sobre modelación de la duración del desempleo tanto a nivel internacional como a nivel nacional. En la tercera sección se presenta un modelo simple de búsqueda de empleo para dar sustento teórico a las estimaciones empíricas de las siguientes secciones. La cuarta se centra en el estimador de Kaplan-Meier, como medida no paramétrica de la duración del desempleo y sus resultados para el caso colombiano. En la quinta se consignan las metodologías econométricas de duración y probabilidad. La sexta describe la información utilizada en los modelos de duración y probabilidad, en particular, la construcción de las variables asociadas a las condiciones del mercado referentes a duración del desempleo y salarios por oficio. En la siguiente sección, se presentan los resultados de los modelos de duración y probabilidad.

Finalmente, en la última sección se muestran los resultados de las diferentes duraciones del desempleo por perfiles de individuo y tipo modelo y la incidencia del desempleo en Colombia.

II. Revisión de la Literatura (con énfasis en métodos de estimación)

Uno de los primeros trabajos sobre la duración del desempleo es el de Ehrenberg y Oaxaca (1976) en éste se realiza una primera aproximación a la explicación de la duración del desempleo. Los datos que se utilizan en el estudio provienen del National Longitudinal Survey (NLS). Este trabajo introduce a la explicación de la duración del desempleo los fundamentos de las teorías de búsqueda de empleo y enfatiza en la importancia de controlar por la variable edad.

En 1979 se publica el primer artículo que explica cómo se debe modelar la duración del desempleo superando los problemas de censura y heterogeneidad presentados en trabajos anteriores. Lancaster (1979) estudia los diferentes problemas que se presentan cuando se interpreta el efecto de las diferencias entre los individuos que buscan empleo y el tiempo que tardan en conseguirlo. El estudio se desarrolla basándose en las teorías de búsqueda de empleo. Los datos utilizados son el resultado de encuestas realizadas sobre una muestra de trabajadores británicos no calificados. Adicionalmente, Lancaster define una lista de variables que deben ser consideradas en la explicación de la duración del desempleo. Estas variables son: i) Tiempo desempleado: esta variable se refiere al tiempo transcurrido entre el momento en que el individuo perdió el último empleo y el momento de la encuesta, ii) empleo reciente, la pregunta se refiere a si el individuo ha encontrado empleo recientemente, es necesario por tanto tener observaciones de individuos que hayan terminado su período de desempleo dentro de la muestra. Es decir, se tienen en cuenta aquellos individuos que estaban en desempleo pero que durante el período de estudio encontraron empleo, iii) edad: esta variable permite explicar muchas diferencias entre los individuos, tales como, el número de ofertas recibidas, iv) desempleo, definido como el porcentaje de la población que está desempleada en la zona donde vive el individuo entrevistado, tal como lo sugiere Lancaster esta variable recoge dos efectos muy cercanos, probabilidad de conseguir empleo y expectativas de conseguirlo. En la sección V se presentará una nueva propuesta para medir expectativas de desempleo, v) Reemplazo: esta variable mide todos los otros ingresos que puede tener una persona, determinantes fundamentales para que el individuo realice un mayor o menor esfuerzo en su búsqueda de empleo. Esta variable es proporcional al salario de reserva de los individuos.

Meyer (1990) muestra, tanto de manera teórica como empírica, que el seguro de desempleo tiene un efecto positivo sobre la duración de éste (un mayor seguro de desempleo conlleva una mayor duración del desempleo). Adicionalmente, introduce el uso de análisis semiparamétrico para construir indicadores que permitan diagnosticar la existencia de variables omitidas. La muestra es conformada por hombres en 12 Estados para el período comprendido entre 1978 y 1983. La fuente es Continuous Wage and Benefit History (CWBH).

Pollman-Schult y Büchel (2005), siguen la línea de Meyer (1990), pero no solo se refiere al seguro de desempleo sino también a una medida más amplia que denominan beneficios del desempleo. Los autores sostienen que mayores beneficios del desempleo aumentarán la duración del mismo pero mejorarán el *match* que se obtenga al final del período de desempleo. A través de estimador Kaplan-Meier y de modelos de regresión Hazard llegan a la conclusión de que las

personas que no reciben beneficios del desempleo efectivamente tienen menor duración de éste pero también tienen mayores tasas de transición a sobre-educarse. Adicionalmente, encuentra que el hecho de recibir beneficios de desempleo no explica la transición a un empleo más calificado. Los datos son tomados del West German Life History Study, recolectado en 1998 para cohortes de individuos nacidos entre 1964 y 1971.

Los trabajos más recientes han desarrollado metodologías más convenientes para el análisis de duración del desempleo. Por ejemplo Fitzenberger y Wilke (2005) desarrollan el método de regresión cuantílica con datos censurados para análisis de duración con aplicaciones a la duración del desempleo, su aplicación se centra en información del oeste de Alemania. Sin embargo, este método aún no permite incluir covariables variantes en el tiempo y no presenta solución a posibles problemas de heterogeneidad no observada. Es de señalar que, este trabajo retoma los desarrollos de Koenker y Billias (2001).

En general, un problema que se enfrenta al trabajar datos de duración del desempleo provenientes de encuestas es que la no respuesta a éstas puede causar diferentes tipos de sesgo en los resultados obtenidos. Van der Berg y Dolton (2006) desarrollan una metodología para identificar entre dos posibles explicaciones del sesgo en los resultados, el primero de selectividad por determinantes no observados de la duración del desempleo y el segundo un efecto causal del motivo de salida del último trabajo y por consiguiente, la no respuesta de la encuesta.

De acuerdo con Wikert y Wilke (2008) los estimadores más utilizados no, necesariamente, son aplicables dada la naturaleza del proceso generador de los datos recolectados y puede causar diferentes formas de censura. Para solucionar este problema se propone una serie de estimadores no paramétricos tipo Kaplan-Meier para las tazas Hazard condicionales y para la función cuantílica condicional. Con este nuevo estimador se superan los problemas que se presentan al emplear los estimadores tradicionales.

Dentro de los trabajos más relevantes para Colombia está el de Tenjo y Ribero (1998) donde además de la duración del desempleo se trabaja la participación laboral. Este estudio utiliza modelación bajo distribución Weibull y realiza estimaciones por género, controlando por estado civil. Emplea datos de la Encuesta Nacional de Hogares para el mes de junio para los años 1988, 1994 y 1996. Encuentra una relación directamente proporcional entre el ingreso familiar y la duración del desempleo de las personas solteras e inversa con la duración en el caso de las personas casadas. Núñez y Bernal (1998) incluyen una sección dedicada a la duración del desempleo en el que utiliza datos del mes de junio de 1988, 1992 y 1996 de la Encuesta Nacional de Hogares. Martínez (2003) hace un estudio sobre duración de empleo y desempleo con datos de la Encuesta Nacional de Hogares.

Castellar y Uribe (2003) llevan a cabo un estudio para Colombia utilizando también modelos paramétricos de duración tipo Weibull. Emplean datos de la encuesta nacional de hogares correspondientes al segundo trimestre de los años 1988, 1992, 1994, 1996 y 1998. El estudio encuentra que la duración promedio del desempleo es de 42 semanas considerando como determinantes de la duración del desempleo las variables ser jefe de hogar (relación negativa con

la duración), ingresos no laborales (a menores ingresos no laborales hay una menor duración del desempleo), género (los hombres tienen menor duración), educación (efecto negativo y no lineal sobre la duración del desempleo), experiencia y dispersión salarial.

Viáfara y Uribe (2008) realizan un análisis no paramétrico para Colombia empleando el estimador Kaplan-Meier. Este trabajo se centra en los canales de búsqueda. Adicionalmente, presentan una amplia revisión de la literatura en cuanto a modelos de Duración y estudios de canales de búsqueda para el caso Colombiano.

III. Un modelo simple de búsqueda de empleo

El problema de búsqueda de empleo ha sido extensamente estudiado en la literatura económica. En dicha literatura se reconoce que la información no es un bien gratuito y que es rentable desde el punto de vista privado¹, dedicar recursos a adquirir información en vez de aceptar el primer trabajo disponible. Una estrategia de búsqueda consiste en fijar un salario de reserva, por debajo del cual se rechaza cualquier oferta de trabajo. Dicho salario de reserva se determina de tal manera que maximice los beneficios netos futuros privados del proceso de búsqueda. Un resumen sencillo de lo que sería la estrategia es el siguiente.

Sea $g(w)$ la función de densidad que describe la percepción del individuo de lo que es la distribución de ofertas salariales en un momento determinado. $g: (\mu, \sigma)$.

Para su búsqueda de empleo, el trabajador fija un salario de reserva w_r tal que

Si $w \geq w_r$ acepta la oferta

Si $w < w_r$ rechaza la oferta

Dado lo anterior, la probabilidad de que una oferta sea rechazada estaría dada por:

$$Prob(w < w_r) = G(w_r) = \int_0^{w_r} g(w)dw \quad (1)$$

La probabilidad de que la oferta sea aceptada estaría dada por:

$$Prob(w \geq w_r) = 1 - G(w_r) \quad (2)$$

Si las ofertas son un proceso aleatorio, el valor esperado del número esperado de ofertas que se recibirían antes de que una sea aceptable está dado por:

$$D = E(L) = \frac{1}{1-G(w_r)} \quad (3)$$

Si se hace el supuesto de que el trabajador recibe una oferta por período, D se puede interpretar como la duración esperada de la búsqueda. Si este supuesto no se cumple, entonces la duración de la búsqueda sería una función de D .

¹ En general se reconoce que hay fallas de mercado que hacen que procesos de búsqueda (inversión en información) sin la adecuada infraestructura institucional no sean socialmente óptimos. Esto se debe a que debido a las externalidades existentes (búsquedas más prolongadas permitan una mejor asignación de recursos, menor rotación de la fuerza de trabajo y una mayor productividad) y al hecho de que la información, una vez obtenida, se convierte en un bien público, las decisiones privadas generalmente resultan en niveles de inversión inferiores a lo socialmente óptimo.

El salario esperado w_e que obtendría el trabajador una vez se reciba una oferta aceptable se puede expresar de la siguiente forma:

$$w_e = E(w|w \geq w_r) = \frac{\int_{w_r}^{\infty} wg(w)dw}{1-G(w_r)} \quad (4)$$

Determinación del Salario de Reserva (w_r):

Una forma sencilla de plantear el problema es suponer que el trabajador fija su salario de reserva de tal manera que maximiza los *beneficios netos* futuros de la búsqueda (diferencia entre beneficios y costos), medidos en términos de valor presente.

Los procesos de búsqueda tienen costos directos y de oportunidad. Los costos directos son los asociados con la búsqueda misma, como la adquisición de información (consulta de avisos, bases de datos, etc.), la generación de ofertas (envío hojas de vida y otras cosas), etc. Los costos de oportunidad están asociados con los ingresos que deja de percibir el trabajador por dedicar su tiempo a la búsqueda en vez de trabajar (el salario que deja de percibir por período). En general, los costos directos se pueden considerar como fijos (por período), pero el costo de oportunidad crece con la duración de la búsqueda. Si suponemos que se recibe una oferta por período de tiempo, podemos expresar los costos (C) como una función de la duración misma de la búsqueda:

$$C = FD + h(D) = C(w_r) \quad \text{y} \quad \frac{\partial C}{\partial w_r} > 0 \quad (5)$$

donde F representa el costo fijo directo por período, h es el costo de oportunidad (que es función de la duración de la búsqueda)

Los beneficios (Y) de la búsqueda están asociados con los ingresos que recibe el trabajador y se pueden medir como una función de w_e .

$$Y = Y(W_r); \quad \text{con} \quad \frac{\partial Y}{\partial w_r} > 0 \quad \& \quad \frac{\partial^2 Y}{\partial w_r^2} < 0 \quad (6)$$

Lo anterior implica que el trabajador determina su salario de reserva óptimo maximizando el valor presente del beneficio neto ($V(B)$) de la búsqueda, dado por la siguiente función:

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

$$V(B) = \frac{1}{r} [Y(w_r)(e^{-rD} - e^{-rN}) - C(w_r)(1 - e^{-rD})] \quad (7)$$

El valor óptimo del salario de reserva (w_r^*) se obtiene resolviendo la siguiente ecuación para w_r :

$$\frac{\partial V(B)}{\partial w_r} = 0 \quad \rightarrow \quad w_r^* = k(g, r, N, X) \quad (8)$$

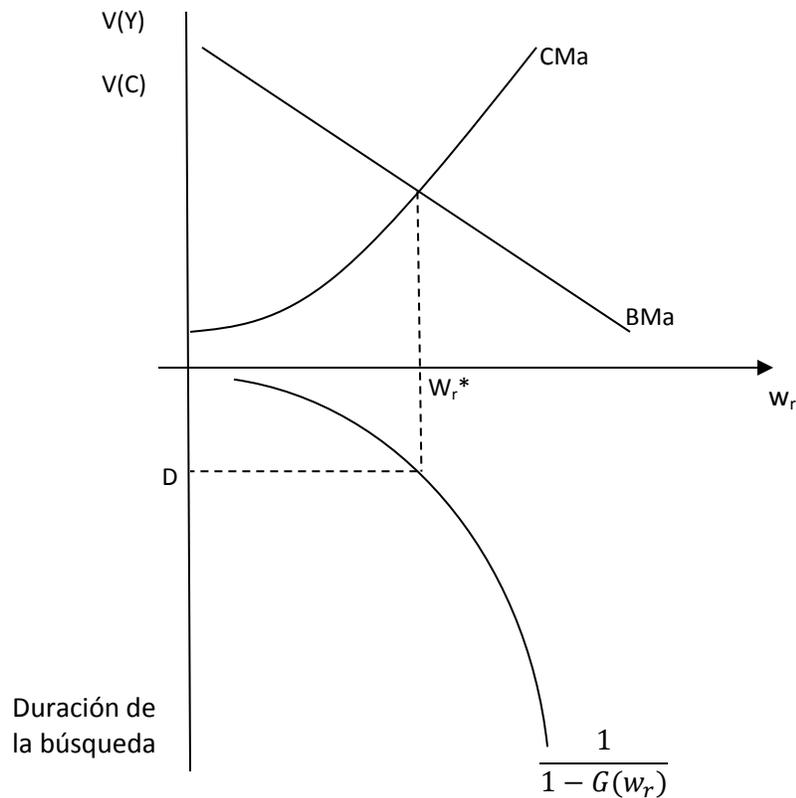
lo que es equivalente a igualar el beneficio marginal con el costo marginal de la búsqueda. La ecuación (8) indica que el salario de reserva óptimo es una función de la distribución de ofertas (g), la tasa de descuento, el horizonte para la toma de decisiones², N (tiempo relevante para efectos de la planeación de la búsqueda y duración del empleo resultante) y un vector de otras variables X .

Una vez conocido w_r^* se puede obtener el valor de D reemplazando en la ecuación (3). De la misma manera, la duración esperada de la búsqueda sería:

$$D = d(g, r, N, X) \quad (9)$$

El Gráfico 1 ilustra este procedimiento.

Gráfico 1



² En modelos más complejos el tiempo esperado de duración del empleo N puede hacerse una función de los resultados de la búsqueda.

La intersección entre el beneficio marginal de la búsqueda (B_{ma}) y el costo marginal de la misma (C_{ma}) determinan el nivel óptimo del salario de reserva (w_r^*). Una vez conocido dicho salario, se remplace en la función de duración (panel inferior del gráfico) para obtener la duración esperada.

Algunas extensiones:

- a. μ se puede hacer función de variables de capital humano y otras que determinan los niveles de ingresos (por ejemplo, sexo, etc.). Esta es una forma sencilla de incluir la diversidad de los mercados laborales. La desviación estándar (σ) se puede suponer constante, pero también se puede hacer depender de algunas características de las personas para captar diferencias en los mercados laborales. Determinar cuáles son las variables responsables por la variabilidad de σ puede ser un problema empírico.
- b. El problema de la tasa de descuento. La tasa de descuento es una medida del costo intertemporal de los recursos necesarios para financiar la búsqueda de empleo. En situaciones donde los mercados funcionan adecuadamente, dicha tasa sería la tasa de interés de equilibrio de largo plazo. Sin embargo, en el caso de la financiación del desempleo el acceso a recursos financieros (en ausencia de intervención estatal) es muy limitado o nulo. Esto constituye una falla de mercado (mercados incompletos) y significa que la financiación de la búsqueda de empleo depende fundamentalmente de los recursos del individuo. Una forma de captar esto es haciendo la tasa de descuento una función inversa de la riqueza de los individuos o, de los ingresos familiares. Esto llevaría a que, otras cosas iguales, la duración de la búsqueda de empleo aumente con los niveles de riqueza o de ingreso familiar y tendría la implicación de que para algunos (muchos?) individuos la *inversión* en información sobre mercado laboral será inferior a lo socialmente óptimo.

Por otro lado, se sabe que la efectividad de los procesos de búsqueda depende en gran medida de la efectividad de los canales de búsqueda y que entre estos, las redes de contacto personales (amigos, conocidos, etc.) son posiblemente los más importantes. Posiblemente la calidad de estos contactos está relacionada con los niveles de riqueza o de ingreso familiar y por lo tanto, las oportunidades de empleo son una función creciente de dicha riqueza.

De esta forma el ingreso familiar juega dos papeles: por un lado representa las posibilidades de los individuos en el proceso de búsqueda y en ese sentido está asociado con las aspiraciones (nivel del salario de reserva); por el otro es una *proxy* de la efectividad de los canales de búsqueda y en ese sentido está asociado con las oportunidades. Estos dos efectos, aspiraciones y oportunidades, operan en sentido contrario en la determinación de la duración de la búsqueda de empleo.

- c. Equivocaciones en la distribución de ofertas salariales. La fijación del salario de reserva depende de la percepción de los individuos sobre lo que es la distribución de ofertas salariales en el mercado laboral. Sin embargo, dada su posición de trabajadores, la

información disponible sobre las condiciones laborales fácilmente puede ser equivocada. Los trabajadores aprenden a cerca del mercado a través de los procesos de búsqueda de trabajo. Esto se puede ver más claramente de la siguiente manera:

Sea un trabajador que asume que la distribución de ofertas salariales es $g_1: (\mu_1, \sigma)$, cuando en realidad es $g_2: (\mu_2, \sigma)$, donde $\mu_1 > \mu_2$. Es decir, el trabajador sobrevalua el verdadero valor de la media de la distribución de ofertas salariales. A partir de la distribución g_1 el trabajador escoge un salario de reserva w_{r1} , lo cual implica que en promedio la duración de la búsqueda sería:

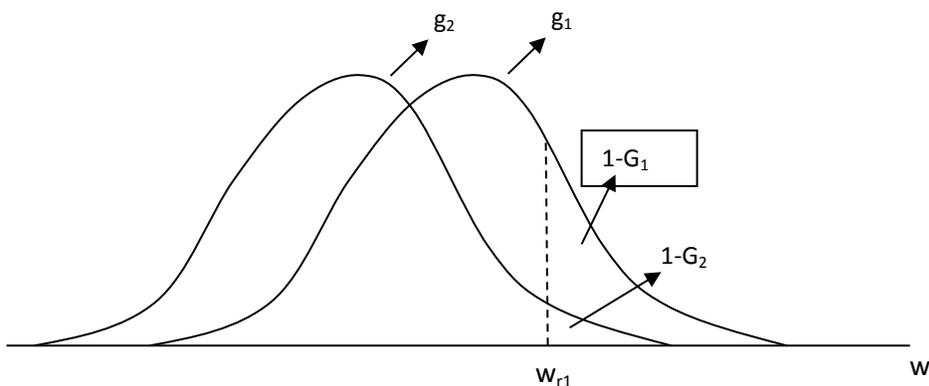
$$D_1 = \frac{1}{1-G_1}$$

Sin embargo, dado que la verdadera distribución es g_2 , la duración de la búsqueda sería:

$$D_2 = \frac{1}{1-G_2} > D_1$$

En este caso el trabajador observa dos cosas: primero que la duración de su búsqueda se prolonga por más allá de lo esperado. Segundo, que después de un tiempo K la serie de ofertas recibidas $\{w_1, w_2, \dots, w_k\}$ no tiene la estructura distributiva que generaría la distribución g_1 . En general lo que observaría es que la proporción de ofertas laborales bajas tendría un peso mayor de que se debería esperar si dichas ofertas fueran generadas por g_1 .

Gráfico 2



En general el trabajador se ve obligado a revisar sus estimativos y trabajar con una distribución más cercana a g_2 , con el agravante que ya ha usado parte de los recursos disponibles para la búsqueda de empleo.

Cuando el error es el contrario (subestimar la media de la distribución de ofertas) el trabajador encontrará un empleo aceptable más rápidamente de lo esperado. En este caso lo más probable es que el trabajador acepte el empleo pero no dure en él la cantidad de tiempo que esperaba.

IV. Estimación No paramétrica de la duración del desempleo

1. Estimador de Kaplan - Meier

La información utilizada para el ejercicio de estimación de la duración del desempleo, como ya se mencionó, proviene de la Gran Encuesta Integrada de Hogares, tercer trimestre de 2010 (Agosto, Septiembre y Octubre). Dicha base de datos contiene información sobre las características personales como tiempo de desempleo, sexo, edad, educación, estado civil, posición en el hogar e ingresos del resto de la familia, todas ellas variables relevantes en la explicación de la duración del desempleo (ver revisión bibliográfica). Sin embargo, la variable duración (tiempo que llevan los desempleados buscando trabajo) es una variable censurada, en el sentido de que presenta el tiempo que llevan los trabajadores buscando empleo al momento de la encuesta, pero no se sabe cuánto tiempo más van a permanecer en esta condición.

Esta censura genera problemas porque introduce sesgos (subvalúa) los indicadores sintéticos de duración del desempleo (como la media y la mediana). Nuestra revisión de literatura a nivel nacional indica que ninguno de los estudios anteriores hechos en Colombia tienen en cuenta este problema y por lo tanto, tampoco presentan alternativas de solución. Aún más, en muchas publicaciones oficiales y no oficiales en las que se usan medidas de duración de desempleo se presentan promedios, medianas y otras medidas de la duración del desempleo basadas en la información existente sin ningún tipo de corrección.

En este documento se propone una metodología para corregir el sesgo de censura, consistente en incluir en la base de datos el tiempo que les tomó conseguir el empleo a los trabajadores que han obtenido el empleo actual recientemente. Es decir, se conforma una variable que especifica que individuos terminaron la búsqueda y cuáles de ellos continúan en ésta. Esto permite tener información sobre la duración completa del desempleo y posibilita la construcción de estimativos consistentes sobre la duración de éste.

Para la medición de la duración del desempleo se utilizaron modelos paramétricos y no paramétricos. Los no paramétricos incluyen la estimación de funciones de supervivencia a través del modelo de Kaplan y Meier, mientras que los paramétricos implican la estimación de modelos de duración ATF (Accelerated Time Failure) con el supuesto de que el error sigue una distribución Weibull.

Metodología No-Paramétrica de la duración del desempleo:

Una primera aproximación a la duración del desempleo se lleva a cabo a través de estadísticas no paramétricas y en particular mediante el estimador de Kaplan-Meier de la función de sobrevivencia, Hosmer et al. 2008, Klein et al. (2003) y Allison (2004)³.

³ Véase, Anexo 1.

2. Resultados del análisis no paramétrico

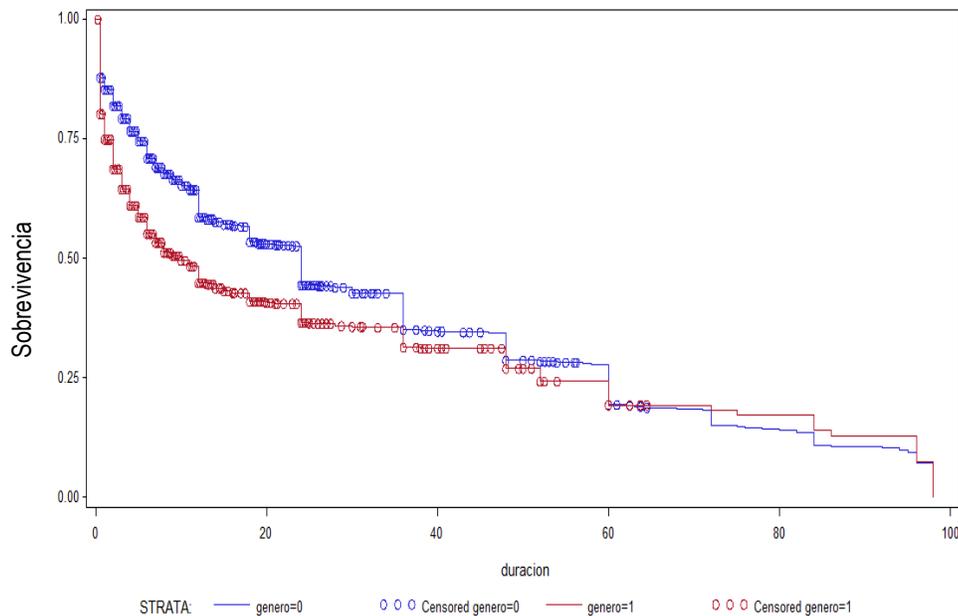
A continuación se presentan los resultados del análisis de duración del desempleo considerando la Gran Encuesta Integrada de Hogares para los meses de julio, agosto y septiembre de 2010. La muestra objeto de estudio se conforma con base en la información de cada uno de estos meses, de tal forma que la variable censura se construye de la siguiente manera:

- La información de julio considera a los individuos no censurados como aquellos individuos que tienen un mes o menos de estar ocupados.
- La información de Agosto considera a los individuos no censurados como aquellos individuos que tienen dos meses o menos de estar ocupados.
- La información de septiembre considera a los individuos no censurados como aquellos individuos que tienen tres meses o menos de estar ocupados.

Así, La muestra objeto del trabajo reúne la información del tercer trimestre con la construcción de la variable censura bajo el esquema ya explicado.

Análisis de la duración del desempleo por género

Gráfico 3



En este caso, la caracterización se realiza por la variable género donde género=1 si la observación corresponde a un hombre y 0 en caso de corresponder a una mujer. Como se deriva del Gráfico 3, la duración del desempleo es menor para los hombres. En particular el 50% de los hombres han conseguido empleo en menos de un año en tanto que, siendo mujer el tiempo requerido para alcanzar el 50% es de 24 meses. La prueba estadística de Wilcoxon de diferencia de medias por género es igual a 0.0001, lo que muestra que el comportamiento entre los dos grupos es estadísticamente diferente

Un análisis similar se hizo para estudiar las diferencias en la duración entre personas con distintas características: (i) jefe de hogar versus no jefe de hogar, (ii) con o sin menores a cargo, (iii) por grupos de edad, (iv) con o sin mayores a cargo, (v) por grados aprobados de educación, (vi) por niveles de ingresos familiares y (vii) por estado civil. Los resultados de estos ejercicios aparecen resumidos en el cuadro 1 considerando la muestra en su totalidad y las sub muestras conformadas por hombres y mujeres.

La variable jefe de hogar, (Jhogar): considera Jhogar=1 si la observación corresponde a un individuo que se reporta como jefe del hogar y 0 en caso contrario. Se observa que la duración del desempleo es menor en el caso de jefes de hogar. En particular, el 50% de los hombres han conseguido empleo a los 5 meses en tanto que, las mujeres con similar condición gastarían, para lograrlo, 18 meses. Las pruebas estadísticas de Wilcoxon, en cada caso, indican diferencia estadística en la duración.

Variable menores a cargo (menores_a_cargo): dicha variable indica si existen niños menores de 7 años en el hogar, donde 1 significa que la observación corresponde a un individuo en cuyo hogar existen menores a cargo y 0 en caso contrario. La menor duración del desempleo se reporta en el caso de hombres con menores a cargo. Como se observa en el Cuadro 1, esta variable no marca diferencia en el caso de las mujeres. Al considerar el total y la sub muestra conformada por hombres se tiene diferencia estadística de la duración. Resultado que no se alcanza en el caso de la muestra conformada por mujeres.

Variable edad: a partir de ésta se conforman 5 grupos de individuos. Así se tiene edad_1 si el individuo es menor de 23 años y 0 en caso contrario, edad_2 si el individuo pertenece al intervalo de edad (22,29] y 0 en caso contrario, edad_3 si el individuo pertenece al intervalo de edad (29, 40] y 0 en caso contrario, edad_4 si el individuo pertenece al intervalo de edad (40,51] y 0 en caso contrario y edad_5 si el individuo es mayor de 51. Así, en general, en los cuantiles 50% y 75%, se observa que a medida que aumenta la edad aumenta la duración del desempleo. Las pruebas estadísticas de diferencia de la duración muestran que el comportamiento entre los grupos es estadísticamente diferente.

Variable mayores a cargo (mayores_a_cargo): esta variable mide si existen en el hogar estudiantes entre 6 y 22 años de edad, (6,22]. La variable mayores_a_cargo=1 indica que la observación corresponde a un individuo en cuyo hogar existen estudiantes con edad en el intervalo

referenciado y 0 en caso contrario. Se encuentra una relación directa entre los individuos con personas mayores a cargo y la duración del desempleo. Las pruebas estadísticas de diferencia de la duración, en este caso, sugieren que el comportamiento considerando el total y la submuestra hombres es estadísticamente diferente.

Variable años de educación aprobados⁴ a partir de la cual se generan 3 variables *dummy*: *aescol1*, *aescol2*, y *aescol3* asociadas a menos de 6 años aprobados, entre 6 y 11 años aprobados y más de 11 años aprobados, respectivamente. En este caso, se observa, para el cuantil 50%, de manera general, que la duración del desempleo aumenta en la medida en que se tiene mayor nivel de escolaridad. Las pruebas estadísticas de diferencia de la duración muestran que el comportamiento entre los grupos es estadísticamente diferente.

Variable ingreso familiar total real neto⁵: a partir de esta variable se conforman 8 grupos de individuos por ingreso. Así se generan 8 variables *dummy* desde *ingfamtotr_1* hasta *ingfamtotr_8* que corresponden a los siguientes intervalos de ingreso: menor a \$200.000, (\$200.000, \$500.000], (\$500.000, \$888.827], (\$888.827, \$1.593.026], (\$1.593.026, \$2.676.212], (\$2.676.212, \$3.768.786], (\$3.768.786, \$6.992.029] y mayores a \$6.992.029. Es de señalar que, en general, no se observa un patrón de comportamiento claro entre los ingresos y la duración del desempleo. Las pruebas estadísticas de diferencia de la duración permiten concluir que el comportamiento entre los grupos es estadísticamente diferente.

Variable estado civil, (*estcivil*): se tiene que *estcivil*=1 si la observación corresponde a una persona casada (o en unión) y 0 en caso contrario. Se observa, en el caso de la submuestra conformada por hombres que la duración del desempleo es menor siendo casado, el 50% de los hombres casados salen del estado de desempleo en el quinto mes. Contrariamente, en el caso de las mujeres, la condición de estar casada aumenta la duración del desempleo y el 50% de ellas requieren de 24 meses para abandonar el estado de desempleo. Las pruebas estadísticas de diferencia de muestras reportan que el comportamiento entre los dos grupos es estadísticamente diferente.

⁴ La construcción de la variable años de educación aprobados (*aeduc*) se presenta en el Anexo 1.

⁵ Sin considerar el ingreso del individuo cuando éste es un individuo ocupado.

Cuadro 1

Características		Total				p-value Test	Hombres				p-value Test	Mujeres			
		Cuantiles			p-value Test		Cuantiles			p-value Test		Cuantiles			p-value Test
		25%	50%	75%			25%	50%	75%			25%	50%	75%	
Género	Mujer	5	24	60	0.0001										
	Hombre	1	10	52											
Jefe de Hogar	No	4	24	60	0.0001	2	12	52	0.0001	6	24	60	0.0001		
	Si	1	9	48		0.5	5	48		3	18	60			
Menores a Cargo	No	3	22	60	0.0001	2	12	60	0.0001	5	24	60	0.9896		
	Si	2	12	48		0.5	5	36		5	24	57			
Edad	<23	3	12	36	0.0010	2	12	36	0.0001	5	18	36	0.0228		
	(22,29]	3	13	42		1	7	36		5	18	48			
	(29,40]	2	12	52		0.5	5	48		5	24	60			
	(40,51]	2	24	60		1	8	52		5	24	64			
	>51	5	36	86		3	36	86		7	36	72			
Mayores a Cargo	No	3	15	58	0.0022	1	9	52	0.0003	5	24	60	0.8159		
	Si	3	18	60		2	11	60		5	24	60			
Grados Aprobados	<6	1	10	48	0.0001	1	5	36	0.0001	3	18	60	0.0001		
	(5,11]	3	15	48		1	8	48		5	24	58			
	>11	6	36	60		4	36	96		3	36	60			
Ingreso Neto (miles)	<200	0.5	5	24	0.0001	0.5	3	12	0.0001	1	8	48	0.0001		
	(200,500]	3	18	60		1	9	60		6	24	60			
	(500,888]	4	24	60		2	12	52		7	24	60			
	(888,1593]	4	24	60		2	12	60		6	24	60			
	(1593,2676]	4	24	60		2	18	96		6	24	60			
	(2676,3768]	4	24	72		3	24	84		7	36	60			
	(3768,6992]	5	36	84		6	48	.		5	24	60			
	>6992	6	48	.		8	.	.		4	48	.			
Estado Civil	No Casado	3	18	60	0.0001	3	15	60	0.0001	4	22	60	0.0001		
	Casado	2	14	52		0.5	5	48		6	24	60			

VI. Modelación econométrica

El análisis econométrico se centra en dos aspectos fundamentales del desempleo en Colombia. El primero, su duración. Con tal propósito se estima el valor esperado de la duración del desempleo a partir de un modelo de regresión paramétrica de duración, donde el logaritmo de la duración se explica mediante un conjunto de variables comúnmente utilizadas en la literatura y el segundo, enmarcado en su probabilidad. Así, con base en modelación Probit y un conjunto de variables explicativas se estudia la probabilidad de que un individuo esté en situación de desempleo.

1. Modelo de duración del desempleo:

A. Conceptos básicos y conformación de la muestra:

Se define la variable aleatoria T como la duración del desempleo, la cual mide el tiempo transcurrido en semanas desde que el individuo abandonó su último empleo y entró al estado de desempleo. La estimación de dicha duración se lleva a cabo mediante un modelo de regresión paramétrica. En particular, se utiliza el modelo *Accelerated Failure Time* o modelo AFT.

El modelo de regresión AFT es especificado a partir de la ecuación (23)

$$\ln(t) = X'\beta + u \quad (23)$$

De tal forma que, diferentes distribuciones del término de perturbación u , como lo son las distribuciones weibull o lognormal, conducen a diferentes modelos AFT. Así, la duración puede expresarse a partir de la ecuación (24)

$$t = \exp(X'\beta)v \quad (24)$$

$$v = e^u \quad (25)$$

Considerando este modelo, existen diferentes formas de llevar a cabo el pronóstico de la duración del evento bajo consideración⁶. En particular, el pronóstico se puede alcanzar a través de la siguiente ecuación:

$$\hat{t}_i = \exp(X_i'\beta + E[u_i])$$

Donde $E[u_i] = \frac{\Gamma(1)}{\hat{\rho}}$ siendo ρ el parámetro de escala de la distribución Weibull y $\Gamma(1)$

la distribución Gamma evaluada en el valor 1.

⁶ Véase, Cleves (2008).

2. Modelo de probabilidad:

El modelo Probit es un caso particular de los modelos de selección discreta donde la variable dependiente y_i es binaria, solo toma dos valores 0 si el individuo está empleado y 1 si este evento no ocurre⁷. Existe un proceso generador de respuesta y_i^* el cual es no observado en la realidad y está definido como la relación siguiente:

$$y_i^* = x_i\beta + u_i \quad (28)$$

$$E[y_i^*] = x_i\beta$$

Lo que se observa en cambio es la variable binaria y_i

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (29)$$

donde $E[y_i] \neq x_i\beta$.

El objetivo es determinar la probabilidad de ocurrencia de un evento condicionado a las características de los individuos de la muestra. La probabilidad de que el evento ocurra está definida como

$$P_i = \text{Prob}(y_i = 1) \quad (30)$$

$$\text{Prob}(y_i = 1) = \text{Prob}(u_i > -x_i\beta)$$

$$= 1 - F_i(-x_i\beta)$$

$$P_i = F_i(x_i\beta)$$

$$X_i = \{\text{características del individuo } i\}$$

$$\beta \rightarrow \text{"Ponderadores por cada característica"}$$

El modelo probit surge de suponer que $u_i \sim N(0, \sigma^2)$. En particular, en este trabajo se revisa la bondad de ajuste del modelo a los datos observados a través de la técnica de calidad de la predicción. Así, se construye una variable indicadora \hat{y}_i

$$\hat{y}_i = \begin{cases} 1 & \text{si } \hat{P}_i > h \\ 0 & \text{si } \hat{P}_i \leq h \end{cases} \quad (33)$$

Donde $0 < h < 1$ es el umbral tal que si la probabilidad estimada supera el umbral la variable indicadora \hat{y}_i toma el valor de 1. Se construye otra variable indicadora w_i donde toma el valor de 1 si el modelo pronosticó correctamente y toma el valor de 0 de lo contrario

$$w_i = \begin{cases} 1 & \text{si } y_i = \hat{y}_i \\ 0 & \text{si } y_i \neq \hat{y}_i \end{cases} \quad (34)$$

⁷ Véase, Cameron y Trivedi (2005) y Maddala (1997).

La calidad de predicción (Cp.) dado un umbral h es el promedio del número de veces que el modelo pronostica correctamente.

$$Cp|h = \frac{1}{N} \sum_{i=1}^N w_i \quad (35)$$

VII. Conjunto de información relevante en la estimación econométrica

A. El modelo de duración del desempleo se estima a partir del modelo AFT, explicado anteriormente, donde la variable dependiente hace referencia a la duración del desempleo como se define a continuación:

1. Duración del desempleo: hace referencia a la dupla de variables que informan sobre cada individuo la duración de su desempleo y su categorización como individuo censurado o no censurado. Es decir, si el individuo ya completó la duración de su desempleo o no, las variables que conforman este grupo son:

- Duración: la variable indica cuánto tiempo lleva desempleado en caso de no haber conseguido empleo o cuánto tiempo se demoró en conseguir el empleo si el individuo ya cambió de estado, es decir consiguió un empleo.
- Censura: esta variable indica si la observación está en censura o ya culminó su periodo de desempleo

En particular, en este trabajo, el conjunto de individuos que culminaron su periodo de desempleo se conforma con la posición del mes dentro del trimestre, como se explicó anteriormente.

El conjunto de variables explicativas o covariables que conforman la matriz de diseño X , para cada una de las muestras mencionadas, se definen a continuación:

2. Edad: La duración del desempleo puede ser explicada, en parte, por la edad de cada individuo. En principio esta variable es un indicador de la experiencia y del conocimiento que el individuo tenga del mercado laboral, pero también está correlacionada con otra serie de factores que pueden afectar el comportamiento del individuo y los resultados de sus acciones. Muchos de estos aspectos no están linealmente relacionados con la probabilidad de desempleo o con la duración de la búsqueda de empleo. Por esa razón, la edad se incluye en las estimaciones como una forma cuadrática (edad y edad al cuadrado). Esto tiene dos efectos: primero, posiblemente los efectos marginales va a depender de los niveles de dicha variable y (2) los test de significancia usuales no son aplicables. En vez de ellos se deben usar tests conjuntos como la razón de verosimilitud.

3. Educación: (aeduc) la variable años aprobados de educación permite capturar el efecto de la especialización sobre la duración del desempleo. Esta variable es de carácter continuo, que al igual que la edad, se introduce en forma cuadrática a fin de capturar efectos más complejos.

4. Reemplazo: el grupo de variables denominadas de reemplazo son todas aquellas variables que permitan explicar el salario de reserva de un individuo. Así, se incorpora de manera indirecta el efecto del salario de reserva sobre la duración del desempleo. Se denominan variables de reemplazo porque son todos aquellos factores que pueden determinar la urgencia que tiene un individuo de conseguir empleo.

- *Estado Civil (estcivil)*: es muy posible que una persona casada tenga mayores responsabilidades implicando un salario de reserva más bajo. La variable toma valor de 1 si el individuo está casado o vive en unión libre.
- *Jefe de hogar (jhogar)*: al igual que la variable anterior, ser jefe de hogar puede estar relacionado negativamente con el salario de reserva.
- *Número de menores a cargo (FNmenor)*: variable de responsabilidad económica sobre otros miembros de la familia⁸ que afecta el salario de reserva en dos vías contrarias. Puede suceder que, entre más personas a cargo se tengan, mayor es el salario de reserva porque su costo de oportunidad es mayor o puede ser que entre más personas a cargo se tenga, su salario de reserva es menor dado que se tiene más necesidad de trabajar
- Ingreso del resto de la familia, *ing_res_fam*, resta del *Ingreso_familiar_total* el ingreso laboral del individuo ocupado. Podría entenderse como el ingreso del resto de la familia y determina el salario de reserva de cada individuo. Posiblemente es un proxy del costo de los recursos disponibles para financiar la búsqueda de empleo y en ese sentido está asociado con el salario de reserva del individuo. Es de señalar que, dicha variable se considera bajo la transformación logarítmica, *ling_res_fam*.
- *Otros controles*: Género (hombres = 1, mujeres = 0)

B. Adicionalmente, se consideran las condiciones de mercado en lo que respecta a duración del desempleo (*dur_cm*) y al salario por ocupación (*ing_cm*), se considera en el modelo su transformación logarítmica (*ling_cm*). La construcción de estas dos variables se lleva a cabo bajo siguiendo la metodología explicada a continuación:

En la literatura internacional se encuentra la sugerencia de utilizar como variable explicativa del modelo de duración del desempleo a las expectativas de dicha duración. Esta variable es de gran importancia porque permite modelar el umbral en el cual las personas empiezan a ajustar sus

⁸ Véase, Anexo 3 (glosario de variables).

estándares. Así, si un individuo ha superado el tiempo esperado de desempleo entonces aceptará ofertas laborales con mayor facilidad, buscará otras opciones u otros cargos para disminuir el tiempo de búsqueda de empleo. En este artículo se usará como variable *proxy* de las expectativas de duración a las condiciones del mercado respecto a dicha variable.

La construcción de la variable proxy de expectativas se lleva a cabo a través de métodos paramétricos que permiten aproximar una distribución conocida a los datos observados de la duración del desempleo. Así, con base en los cuantiles de la distribución se llega a inferir las condiciones del mercado sobre el tiempo de desempleo para cada una de las ramas de actividad reportadas en la encuesta⁹.

Así, se considera la información correspondiente a duración del desempleo y la ocupación para la muestra de ocupados, con menos de tres meses de ocupación respecto a cada uno de los meses que conforman el tercer trimestre de 2010, Posteriormente, se aproxima para cada una de estas ocupaciones una distribución asociada al tiempo reportado de desempleo. Es de señalar que se tiene información a priori acerca de la distribución que pueden seguir los datos, ésta debe ser positiva y continua, dada la naturaleza del problema. La estadística sugiere¹⁰ como distribuciones posibles las distribuciones gamma, log-normal, Chi-cuadrado, exponencial, F (snedecor) y Weibull. Los parámetros se estiman por el método de máxima verosimilitud, ecuación 37, y una vez se determina la distribución que siguen los datos se construyen las expectativas de la duración del desempleo en los cuantiles de interés.

$$\text{Max}_{\mu, \sigma} \prod_{i=1}^T f(X_i | \mu, \sigma) \quad (37)$$

La selección de la distribución se lleva a cabo a través de una medida de evaluación del ajuste de ésta a los datos. Es decir, se selecciona aquella que presente la menor diferencia entre los datos observados y la distribución estimada (d_i). En particular, como criterio de selección se utiliza el método de error cuadrático medio (ECM)¹¹, ecuación (38).

$$\text{ECM} = \frac{\sum_{i=1}^T d_i^2}{T} \quad (38)$$

⁹ Surge la pregunta de ¿porqué no utilizar el tiempo promedio del desempleo para cada uno de los oficios? y la respuesta es que debido a la alta dispersión presentada, el promedio no daría una medida de tendencia central confiable.

¹⁰ Véase, Evans et al. (2011).

¹¹ Véase, un ejemplo particular en el Anexo 2

Aplicando esta metodología a la duración del desempleo por ocupación se encuentra que las distribuciones que mejor se aproximan al tiempo de la duración del desempleo por ocupación son las distribuciones Gamma y Weibull¹². Para la construcción de la variable de condiciones del mercado referentes al tiempo de duración del desempleo por ocupación se toma el cuantil 75, dado que éste exhibe una mayor concentración de masa de probabilidad. Así, este es el umbral de duración donde el 75% de los individuos de la muestra consiguen empleo.

Se encuentra la distribución Gamma como la mejor distribución que aproxima al salario mensual por ocupación. Para la construcción de la variable de condiciones del mercado en lo referente al salario por ocupación se toma el cuantil 75 dado que éste exhibe una mayor concentración de masa de probabilidad. Un individuo con expectativas elevadas de salario puede ser que se demore más tiempo en conseguir empleo y considere, en algún momento, cambiar de oficio para alcanzar el nivel deseado. Para cada una de las ramas de ocupación de la muestra se exhiben expectativas de salario diferentes que están relacionadas con el perfil necesario en cada ocupación.

Es de resaltar que, los resultados de las condiciones del mercado en lo referente al cuantil 75 tanto de la duración del desempleo (meses) como del salario (Salario medido en pesos deflactado por el índice de precios de cada mes) se construyen por cada ocupación, teniendo en cuenta género, edad¹³ y años de educación¹⁴.

¹² Es de señalar que, la distribución Weibull es un caso particular de la distribución Gamma ante modificaciones en el parámetro de forma.

¹³ Grupo 1: menores a 30 años, Grupo 2: entre 30 y 50 años y Grupo 3: mayores a 50 años.

¹⁴ Grupo 1: menos de 6 años de educación, Grupo 2: entre 6 y 11 años de educación y Grupo 3: más de 11 años de educación.

VIII. Resultados de los modelos econométricos

Las estimaciones econométricas se llevan a cabo con base en la información contenida en la muestra total y en cada una de las sub muestras conformadas por hombres y mujeres, respectivamente. En particular, se conforman 4 matrices de diseño y por consiguiente, cuatro modelos por estimar:

- $X1 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2}\}$
- $X2 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, jhogar, estcivil, FNmenor}\}$
- $X3 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, ling_cm, dur_cm}\}$
- $X4 = \{\text{Género, edad, edad2, ling_res_fam, ling_res_fam2, jhogar, estcivil, FNmenor, ling_cm, dur_cm}\}$

A. Resultados de la estimación de los modelos de duración:

La estimación de los diferentes modelos de duración se lleva a cabo a través del modelo AFT, explicado en la sección VIB. En las distintas estimaciones se considera que el término de perturbación sigue una distribución Weibull. Es de señalar que, la decisión acerca de la distribución se toma con base en pruebas realizadas considerando otras distribuciones¹⁵. En particular, se considera la distribución exponencial, como un caso particular de la Weibull, y se rechaza su existencia mediante un test de multiplicadores de Lagrange.

La estimación de los cuatro modelos se realiza sobre la muestra total y sobre las sub muestras construidas considerando la información asociada a hombres y a mujeres. En particular, al considerar la información total, los diferentes modelos incluyen la variable género. En el caso de las sub muestras dicha variable es excluida.

Los Cuadros 2, 3 y 4 presentan los resultados de la estimación de los cuatro modelos considerando la muestra total y las sub muestras generadas por género, respectivamente.

¹⁵ Entre otras, Lognormal y Exponencial.

Cuadro 2

Muestra Total								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value
Intercept	2.2324	<.0001	2.0611	<.0001	3.2392	<.0001	3.0957	<.0001
genero	-0.7499	<.0001	-0.7016	<.0001	-0.479	<.0001	-0.4466	<.0001
edad	-0.068	<.0001	-0.0402	<.0001	-0.0487	<.0001	-0.024	0.002
edad2	0.0012	<.0001	0.0009	<.0001	0.0008	<.0001	0.0006	<.0001
aeduc	0.0815	<.0001	0.0753	<.0001	0.0516	0.0009	0.0486	0.0016
aeduc2	0.0003	0.7367	0.0001	0.8916	0.0015	0.0869	0.0012	0.161
ling_res_fam	0.0673	0.0006	0.0983	<.0001	0.0678	0.0004	0.0959	<.0001
ling_res_fam2	0.0017	0.2023	-0.0011	0.418	0.0015	0.2602	-0.0011	0.4203
ling_cm					-0.1198	<.0001	-0.1188	<.0001
dur_cm					0.0445	<.0001	0.0414	<.0001
jhogar			-0.5827	<.0001			-0.5467	<.0001
estcivil			-0.1647	<.0001			-0.1441	<.0001
FNmenor			-0.0903	<.0001			-0.0901	<.0001
Scale	1.4444		1.4311		1.4148		1.4037	
Weibull Shape	0.6923		0.6988		0.7068		0.7124	

Cuadro 3

Sub muestra: Hombres								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value
Intercept	2.3275	<.0001	1.7465	<.0001	2.4618	<.0001	1.8416	<.0001
edad	-0.1228	<.0001	-0.0589	<.0001	-0.0917	<.0001	-0.0346	0.0009
edad2	0.0019	<.0001	0.0013	<.0001	0.0013	<.0001	0.0008	<.0001
aeduc	0.0737	0.0008	0.0687	0.0012	0.051	0.0162	0.0497	0.0158
aeduc2	0.0027	0.0325	0.0019	0.1099	0.002	0.0895	0.0013	0.2775
ling_res_fam	-0.0103	0.6978	0.0318	0.22	0.003	0.9082	0.0426	0.0906
ling_res_fam2	0.0069	0.0002	0.0031	0.0916	0.0056	0.0018	0.002	0.2541
ling_cm					-0.0572	0.0007	-0.0478	0.0034
dur_cm					0.0925	<.0001	0.0863	<.0001
jhogar			-0.4059	<.0001			-0.3771	<.0001
estcivil			-0.6686	<.0001			-0.6163	<.0001
FNmenor			-0.1362	<.0001			-0.1337	<.0001
Scale	1.468		1.4335		1.4183		1.3882	
Weibull Shape	0.6812		0.6976		0.7051		0.7203	

Cuadro 4

Sub muestra: Mujeres								
	Modelo 1		Modelo 2		Modelo 3		Modelo 4	
	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value	Coficiente	P-value
Intercept	1.2054	<.0001	1.3754	<.0001	2.7168	<.0001	2.851	<.0001
edad	0.0035	0.7538	-0.0069	0.5532	0.034	0.0023	0.023	0.0506
edad2	0.0002	0.16	0.0004	0.0104	-0.0003	0.08	-0.0001	0.6585
aeduc	0.0775	0.0005	0.0769	0.0005	0.0292	0.1998	0.0301	0.1845
aeduc2	-0.0015	0.1975	-0.0014	0.2246	0.0015	0.2219	0.0015	0.2213
ling_res_fam	0.1548	<.0001	0.1772	<.0001	0.1515	<.0001	0.1709	<.0001
ling_res_fam2	-0.0044	0.0237	-0.0065	0.0009	-0.0043	0.0262	-0.0061	0.0016
ling_cm					-0.1644	<.0001	-0.1601	<.0001
dur_cm					0.0224	<.0001	0.0214	<.0001
jhogar			-0.3566	<.0001			-0.3256	<.0001
estcivil			0.2333	<.0001			0.2196	<.0001
FNmenor			-0.007	0.7942			-0.0117	0.6587
Scale	1.3951		1.3865		1.3772		1.3705	
Weibull Shape	0.7168		0.7212		0.7261		0.7297	

Las estimaciones del cuadro 2 arrojan resultados interesantes. En general se encuentra que los hombres tienen procesos de búsqueda de empleo más cortos que las mujeres. Esto se puede explicar porque las mujeres, al ser las principales responsables por las tareas domésticas y el cuidado de los niños, tienen mayores demandas sobre su tiempo y, por lo tanto, un mayor salario de reserva que los hombres (otras cosas iguales), según lo indica la teoría presentada anteriormente.

También se encuentra que las personas casadas y los jefes de hogar tienen desempleo de más corta duración que el resto de la población. Esto es consistente con la hipótesis de que dichas personas, por ser los principales responsables del sostenimiento del hogar tienen salarios de reserva menores y por lo tanto menores períodos de búsqueda de empleo (son menos exigentes con respecto a las condiciones de trabajo).

La combinación de signos de la variable edad (negativo para el primer término y positivo para el término cuadrático) indica que la duración del desempleo disminuye con la edad hasta cierto punto pero a partir de dicho punto aumenta. La edad crítica (punto donde la duración del desempleo deja de decrecer y comienza a aumentar) está entre 20 y 30 años de edad, dependiendo del modelo que se esté usando¹⁶. Esto en general implica que, otras cosas iguales, la duración del desempleo de los jóvenes es menor que la de las personas mayores.

Los años de educación (aeduc) tienen un efecto cuadrático que aumenta la duración del desempleo. Las personas con bajos niveles de educación tienen períodos de desempleo de baja duración, pero a medida que los niveles educativos aumentan, los

¹⁶ Este punto está dado por los coeficientes de edad y edad² = - edad/(2*edad²)

procesos de búsqueda se hacen más prolongados. Esto se puede deber a que entre mayores sean los niveles educativos, mayor la complejidad de los empleos y por lo tanto mayor la información que los buscadores necesitan obtener antes de aceptar un empleo. Esto hace que los procesos de búsqueda sean más prolongados.

El ingreso del resto de la familia (*ling_res_fam*), como ya se dijo, está relacionado con la disponibilidad de recursos para financiar la búsqueda de empleo¹⁷. En ese sentido, entre mayor sea el ingreso del resto de la familia, mayor será el salario de reserva de la persona y (otras cosas iguales) mayor será la duración de la búsqueda de empleo. Esta relación se conoce como el “*efecto aspiraciones*”. Sin embargo, en un país como Colombia en donde la mayor parte de los trabajadores consiguen empleo a través de contactos, amigos, conocidos, etc., el ingreso del resto de la familia también es un proxy de la calidad de dichas redes. El efecto de la calidad de los contactos es hacer más eficientes los procesos de búsqueda y por lo tanto mejorar las ofertas laborales que el trabajador hace, reduciendo de esta forma la duración esperada del desempleo. Este segundo efecto lo podemos llamar “*efecto oportunidades*”.

Los resultados del Cuadro 2 indican que en general predomina el efecto aspiraciones, aunque los signos negativos de la forma cuadrática del modelo 2 y el modelo 4 son evidencia de que a niveles altos de ingreso el efecto oportunidades puede ser importante.

Es importante mencionar los resultados de las variables que miden las condiciones del mercado laboral como son el salario esperado (*Ling_cm*) y la duración esperada de la búsqueda de empleo (*dur_cm*).

Los resultados del Cuadro 2, modelos 3 y 4, indican que entre mayor sea el salario esperado (medido por lo que recibe los trabajadores que han conseguido empleo recientemente) menor será la duración del desempleo. Esto se puede interpretar en el sentido de que las ocupaciones que, otras cosas iguales, pagan mejores salarios, probablemente son ocupaciones donde hay presiones de demanda y por lo tanto la consecución de empleo en ellas es más fácil. Posiblemente los empleadores hacen mayores inversiones en la generación de información para facilitar el enganche de trabajadores.

La duración esperada del desempleo (*dur_cm*) arroja signo positivo, lo cual indica que en las ocupaciones donde es más difícil conseguir empleo, los trabajadores que buscan dicho empleo están dispuestos a hacer búsquedas más prolongadas.

En conjuntos los resultados de estas dos últimas variables muestran que las condiciones de los mercados (por ocupaciones) son importantes en los procesos de búsqueda de empleo.

¹⁷ En ausencia de seguros de desempleo, los trabajadores dependen de sus propios ingresos y ahorros para financiar la búsqueda de empleo. En general, los mercados financieros no generan mecanismos para financiar los procesos de búsqueda de trabajo (fallas de mercado).

Los resultados de los cuadros 3 y 4 para hombres y mujeres, respectivamente, confirman en general los resultados anteriores, con algunas excepciones que mencionamos a continuación:

Tal vez el principal cambio en los resultados cuando se desagrega por sexo es el efecto del estado civil. Para los hombres esta variable sigue teniendo un efecto negativo, pero para las mujeres es positivo. Este resultado es consistente con la hipótesis ya planteada de que, otras cosas iguales, las mujeres tienen salarios de reserva más altos que los hombres y por eso sus procesos de búsqueda son más exigentes y prolongados. El hecho de que el estar casadas tenga un efecto positivo sobre la duración del desempleo de las mujeres indica precisamente esto.

B. Resultados de la estimación de los modelos de probabilidad:

A continuación se presentan los resultados de la estimación del modelo Probit robusto, el cual está corregido por la heterogeneidad no observada.

La muestra en los modelos de probabilidad está conformada por el total de ocupados y desocupados de la Gran Encuesta Integrada de Hogares, agregación de los meses correspondientes al tercer trimestre de 2010. La variable dependiente es 1 si la persona está desempleada y cero si está ocupado. Es de señalar que, tan solo se consideran las matrices de diseño X_1 y X_2 . Es decir, no se consideran las condiciones de mercado en lo referente a duración y salario.

En los cuadros 5, 6 y 7 se consignan los resultados de la estimación del modelo Probit robusto para la muestra total y las sub muestras construidas mediante la variable género, respectivamente

En la primera columna de cada una de las muestras y de cada uno de los modelos se encuentran los coeficientes estimados, los cuales, dada la no linealidad, no son directamente interpretables. La segunda columna es el *P-value* asociado a los coeficientes estimados.

Cuadro 5

Muestra: Total				
	Modelo 1		Modelo 2	
	Coeficiente	P-value	Coeficiente	P-value
Intercept	0.1165	0.0120	-0.0652	0.1820
genero	-0.2221	0.0000	-0.1633	0.0000
edad	-0.0633	0.0000	-0.0467	0.0000
edad2	0.0006	0.0000	0.0004	0.0000
aeduc	0.0176	0.0000	0.0179	0.0000
aeduc2	-0.0009	0.0000	-0.0009	0.0000
ling_res_fam	0.0395	0.0000	0.0512	0.0000
ling_res_fam	-0.0013	0.0030	-0.0025	0.0000
jhogar			-0.2326	0.0000
estcivil			-0.1683	0.0000
FNmenor			0.0016	0.8080

Cuadro 6

Sub muestra: Hombres				
	Modelo 1		Modelo 2	
	Coeficiente	P-value	Coeficiente	P-value
Intercept	-0.1325	0.0340	-0.4476	0.0000
edad	-0.0704	0.0000	-0.0398	0.0000
edad2	0.0007	0.0000	0.0005	0.0000
aeduc	0.0110	0.1000	0.0119	0.0780
aeduc2	-0.0002	0.6260	-0.0002	0.4710
ling_res_fam	0.0165	0.0570	0.0483	0.0000
ling_res_fam	0.0008	0.2060	-0.0020	0.0010
jhogar			-0.3380	0.0000
estcivil			-0.2581	0.0000
FNmenor			-0.0453	0.0000

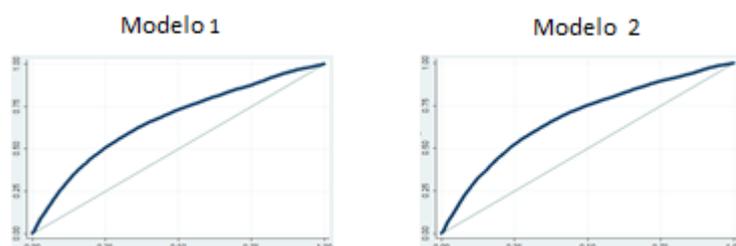
Cuadro 7

Sub muestra: Mujeres				
	Modelo 1		Modelo 2	
	Coeficiente	P-value	Coeficiente	P-value
Intercept	-0.0313	0.6570	-0.1010	0.1640
edad	-0.0440	0.0000	-0.0394	0.0000
edad2	0.0002	0.0000	0.0002	0.0000
aeduc	0.0219	0.0020	0.0228	0.0010
aeduc2	-0.0015	0.0000	-0.0015	0.0000
ling_res_fam	0.0640	0.0000	0.0646	0.0000
ling_res_fam	-0.0034	0.0000	-0.0035	0.0000
jhogar			-0.0718	0.0000
estcivil			-0.0617	0.0000
FNmenor			0.0339	0.0000

El Gráfico 4 muestra la calidad de predicción de los modelos 1 y 2 para el total de la muestra. En general, el modelo 1 pronostica correctamente el 67% de las veces, en tanto que, el segundo modelo lo hace el 68% de las veces. La línea azul representa la calidad de la predicción para cada uno de los umbrales¹⁸. La línea delgada es la calidad de pronóstico de una variable aleatoria con distribución Bernoulli.

Gráfico 4

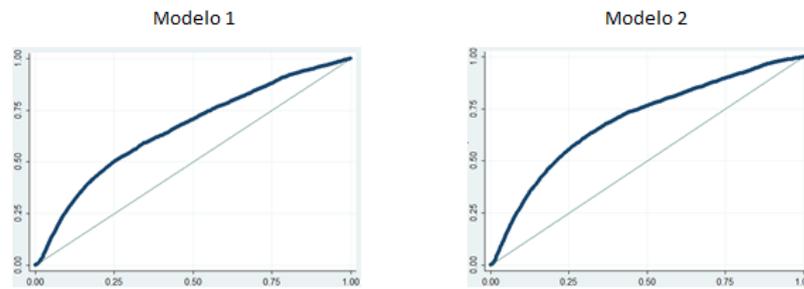
Muestra Total



¹⁸El umbral se define como un valor en el intervalo (0,1) que mide la proporción de aciertos del pronóstico versus lo observado.

Gráfico 5

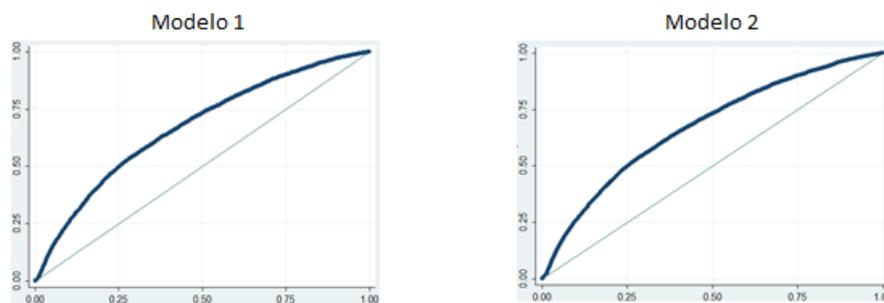
Sub muestra Hombres



El Gráfico 5 muestra la calidad de predicción de los modelos 1 y 2 para la sub muestra conformada por hombres. El modelo 1 pronostica correctamente el 66% de las observaciones, en tanto que, el segundo modelo lo hace para el 69.4% de éstas.

Gráfico 6

Sub muestra Mujeres



El Gráfico 6 muestra la calidad de predicción de los modelos 1 y 2 para la sub muestra conformada por mujeres. Los dos modelos pronostican correctamente el 67% de las observaciones.

IX. Perfiles

A. Duración

A continuación se presentan los pronósticos de la duración del desempleo por perfiles de individuo, considerando los parámetros estimados en cada una de las sub muestras (hombres y mujeres). Es de señalar que, se consideran los promedios de las variables explicativas de cada modelo de duración correspondiente a cada sub muestra particular. Así, el promedio del ingreso familiar total y la variable que miden menores a cargos, son obtenidos sobre cada una de las sub muestras¹⁹ seleccionando por estado civil, jefe de hogar y edad.

El cuadro 8 presenta los resultados de los pronósticos de la duración del desempleo por perfil considerando la sub muestra conformada por hombres. Los cuatro modelos considerados.

En el cuadro 9 se consignan los resultados de la duración del desempleo por perfil considerando la muestra de mujeres.

Como se observa, de manera general, la duración del desempleo es creciente con la edad y la educación. Así, una persona que pierde el empleo tarda más tiempo en conseguir un nuevo empleo entre más edad y mayor educación presente. Igualmente, se concluye que las medidas de duración correspondientes a las mujeres son siempre mayores a las reportadas por los hombres. Es decir, se alcanzan resultados robustos frente a la selección de la muestra y a los diferentes modelos. Es de señalar que, el modelo 4 reporta las menores duraciones del desempleo en las dos sub muestras consideradas.

¹⁹ Se tienen tres sub muestras de datos definidas a partir del tiempo que llevan desempleados en según el mes del tercer trimestre.

Cuadro 8 - Hombres

Jefe de Hogar=0 , Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	2.2	3.3	7.1	16.2	2.0	3.0	6.4	15.4	2.6	4.7	8.7	19.5	4.9	8.2	16.3	36.0
Modelo 2	3.3	4.9	9.5	19.7	4.1	6.2	11.5	23.0	6.4	10.8	18.5	36.0	12.6	19.8	36.0	70.0
Modelo 3	2.3	3.3	5.5	10.4	2.1	2.9	4.7	8.8	2.4	3.5	5.7	9.7	3.7	5.3	9.3	15.1
Modelo 4	3.3	4.8	7.5	13.2	4.1	5.7	8.4	13.8	5.8	8.1	12.2	19.0	9.4	12.9	21.2	31.7
Jefe de Hogar=0 , Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	1.9	3.1	6.8	15.6	1.8	2.9	6.3	15.4	2.5	4.1	8.7	19.3	5.0	8.2	17.4	44.4
Modelo 2	1.4	2.3	4.4	9.3	1.9	2.9	5.3	11.1	3.2	4.9	9.1	17.8	6.5	10.2	19.1	41.6
Modelo 3	2.2	3.0	5.2	9.8	1.7	2.7	4.5	9.0	2.2	3.2	5.7	9.6	3.7	5.2	9.7	18.3
Modelo 4	1.6	2.3	3.6	6.4	1.9	2.7	4.0	7.1	2.8	4.0	6.4	9.8	5.0	6.9	11.6	19.6
Jefe de Hogar=1 , Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	0.6	3.1	5.9	13.2	1.8	2.7	5.3	11.9	2.5	4.0	7.5	16.3	4.9	6.9	14.9	29.0
Modelo 2	0.9	3.4	5.8	11.8	2.9	4.0	7.0	13.4	4.3	6.6	11.4	21.7	8.7	12.3	23.3	40.3
Modelo 3	0.8	3.0	4.5	8.1	1.7	2.4	3.9	6.8	2.1	3.1	4.9	8.7	4.0	4.8	8.6	12.0
Modelo 4	1.1	3.4	4.7	7.9	2.7	3.6	5.4	8.2	3.7	5.2	7.7	12.6	7.2	8.7	14.3	18.5
Jefe de Hogar=1 , Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	1.6	2.8	5.8	12.9	1.7	2.7	5.6	14.4	2.4	3.8	8.0	19.3	5.0	8.0	15.9	46.2
Modelo 2	0.9	1.4	2.7	5.5	1.2	1.9	3.5	7.3	2.0	3.1	5.9	11.8	4.4	6.7	12.1	28.1
Modelo 3	1.9	2.8	4.4	8.1	1.7	2.5	4.1	8.1	2.1	3.0	5.2	10.0	3.6	5.0	8.9	19.6
Modelo 4	1.1	1.5	2.3	4.0	1.3	1.8	2.7	4.6	1.9	2.6	4.2	7.1	3.4	4.6	7.6	14.0

Cuadro 9 – Mujeres

Jefe de Hogar=0, Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.5	13.3	16.5	7.8	11.2	15.5	18.9	9.9	13.7	18.8	22.9	12.2	17.1	23.4	28.8
Modelo 2	6.3	9.1	12.6	15.4	7.5	10.9	14.9	17.9	10.3	14.0	19.1	23.2	13.8	19.3	25.8	31.3
Modelo 3	6.3	7.9	10.9	12.4	8.0	9.7	12.9	12.9	9.2	11.0	14.5	16.1	10.5	12.9	17.0	18.3
Modelo 4	6.0	7.6	10.5	11.8	7.8	9.4	12.5	12.4	9.5	11.2	14.7	16.4	11.6	14.4	18.6	19.8
Jefe de Hogar=0, Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.7	13.5	16.8	8.0	9.9	8.0	9.7	9.6	13.7	18.9	23.2	12.1	17.2	23.6	28.7
Modelo 2	8.1	11.6	16.1	19.8	11.3	13.9	10.1	12.2	12.4	17.5	24.0	29.1	17.3	24.1	32.2	38.9
Modelo 3	6.6	8.7	11.2	12.4	15.6	18.9	12.9	15.6	9.3	11.4	15.1	15.8	11.5	13.2	17.5	16.5
Modelo 4	8.0	10.3	13.4	14.7	19.1	22.8	13.7	16.4	11.8	14.4	19.0	19.8	15.8	18.0	23.4	22.3
Jefe de Hogar=1, Estado Civil=0																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.3	9.6	13.1	16.1	7.9	11.1	15.3	18.3	9.6	13.5	18.5	22.4	12.1	17.1	23.4	28.3
Modelo 2	4.4	6.5	8.9	11.1	5.5	7.7	10.6	12.8	7.0	9.8	13.5	16.5	9.8	13.6	18.3	22.3
Modelo 3	6.9	8.6	10.7	11.7	7.7	9.4	12.4	13.2	8.9	10.9	14.7	16.0	10.5	12.9	17.7	17.9
Modelo 4	5.0	6.0	7.5	8.4	5.5	6.8	8.9	9.5	6.7	8.2	11.0	12.2	8.6	10.4	14.2	14.4
Jefe de Hogar=1, Estado Civil=1																
Edad	25				35				45				55			
Años de Educación	0	5	11	16	0	5	11	16	0	5	11	16	0	5	11	16
Duración del desempleo																
Modelo 1	6.7	9.6	13.3	16.3	8.2	11.3	15.4	18.8	9.7	13.6	18.7	22.6	12.1	17.0	23.3	29.2
Modelo 2	5.7	8.2	11.3	13.9	7.1	9.7	13.3	16.2	8.7	12.3	17.0	20.3	12.5	16.7	22.7	27.0
Modelo 3	7.1	8.0	10.8	12.5	7.8	9.5	13.1	13.0	9.2	10.7	14.6	15.8	11.8	12.4	18.1	14.4
Modelo 4	6.1	7.0	9.4	11.0	6.9	8.4	11.5	11.5	8.4	9.9	13.5	14.5	12.0	12.3	17.8	13.7

B. Probabilidad del desempleo

Los Cuadros 10 y 11 presentan los pronósticos de la probabilidad de estar desempleado dado un conjunto de características o perfiles del individuo de interés, edad, estado civil, años de educación y jefe de hogar. Es de señalar que, las restantes variables explicativas se consideran en el promedio de los individuos que conforman la caracterización particular. En este caso, las probabilidades pronosticadas se basan en los modelos 1 y 2, reportados en los cuadros 6 y 7. El cuadro 10 reporta las probabilidades obtenidas en el caso de hombres, en tanto que el cuadro 11 presenta los resultados para mujeres. Como se destaca, la probabilidad de estar en desempleo es menor entre mayor sea la persona. Dentro de cada edad, en cada uno de los perfiles, los años de educación aumentan la probabilidad de estar en desempleo. En general, la probabilidad de estar desempleado es mayor en el caso de las mujeres.

De los cuadros anteriores se puede concluir que, si bien es cierto, una mayor edad está asociada a una mayor duración del desempleo también está asociada a una menor probabilidad de perder el empleo. Con el fin de concluir al respecto, es necesario construir medidas de incidencia, las cuales se presentan en la siguiente sección.

Cuadro 10 - Hombres

Jefe de Hogar=0, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.228	0.249	0.266	0.276	0.191	0.202	0.215	0.223	0.173	0.184	0.193	0.198	0.172	0.183	0.191	0.200
Modelo 2	0.287	0.300	0.312	0.319	0.264	0.274	0.286	0.293	0.257	0.268	0.279	0.286	0.266	0.279	0.290	0.297
Jefe de Hogar=0, Estado Civil = 1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.237	0.247	0.260	0.271	0.189	0.199	0.212	0.220	0.171	0.179	0.191	0.199	0.179	0.185	0.195	0.201
Modelo 2	0.232	0.242	0.253	0.260	0.213	0.222	0.231	0.238	0.208	0.219	0.228	0.234	0.218	0.229	0.238	0.245
Jefe de Hogar=1, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.191	0.215	0.224	0.232	0.161	0.173	0.178	0.182	0.143	0.156	0.161	0.163	0.152	0.157	0.165	0.167
Modelo 2	0.183	0.220	0.229	0.236	0.184	0.200	0.204	0.208	0.175	0.194	0.199	0.200	0.194	0.202	0.212	0.212
Jefe de Hogar=1, Estado Civil = 1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.221	0.231	0.242	0.253	0.182	0.187	0.197	0.206	0.167	0.174	0.181	0.186	0.171	0.178	0.186	0.191
Modelo 2	0.174	0.182	0.191	0.199	0.160	0.168	0.175	0.181	0.157	0.166	0.173	0.178	0.166	0.175	0.183	0.188

Cuadro 11 - Mujeres

Jefe de Hogar=0, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.335	0.354	0.349	0.322	0.276	0.291	0.285	0.262	0.229	0.243	0.239	0.217	0.198	0.209	0.205	0.187
Modelo 2	0.350	0.368	0.359	0.328	0.289	0.303	0.297	0.272	0.242	0.256	0.252	0.227	0.211	0.222	0.218	0.197
Jefe de Hogar=0, Estado Civil = 1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.338	0.348	0.344	0.321	0.272	0.285	0.280	0.261	0.225	0.237	0.234	0.216	0.195	0.204	0.200	0.185
Modelo 2	0.338	0.346	0.341	0.317	0.271	0.283	0.279	0.260	0.226	0.237	0.233	0.215	0.196	0.205	0.201	0.185
Jefe de Hogar=1, Estado Civil = 0																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.310	0.334	0.336	0.315	0.264	0.281	0.275	0.254	0.230	0.243	0.239	0.215	0.196	0.209	0.207	0.188
Modelo 2	0.316	0.333	0.332	0.309	0.263	0.280	0.274	0.253	0.232	0.243	0.240	0.215	0.197	0.210	0.208	0.189
Jefe de Hogar=1, Estado Civil = 1																
Edad	25				35				45				55			
Años Educación	0	5	11	16	0	5	11	16	0	5	11	16	13	14	15	16
Probabilidad del Desempleo																
Modelo 1	0.318	0.350	0.349	0.328	0.276	0.290	0.286	0.266	0.228	0.243	0.241	0.223	0.198	0.209	0.207	0.191
Modelo 2	0.301	0.336	0.333	0.310	0.263	0.277	0.273	0.254	0.217	0.232	0.229	0.213	0.188	0.200	0.198	0.182

VII Incidencia del desempleo

Una relación muy importante es la que define la tasa de desempleo como el producto de la incidencia por la duración del desempleo:

Tasa de desempleo mensual = duración de desempleo (en meses) * Tasa de incidencia mensual

La tasa de incidencia se puede definir como la proporción de la fuerza de trabajo que entra mensualmente a la situación de desempleo.

Esta relación permite distinguir entre diferentes tipos de desempleo: por ejemplo el desempleo puede ser alto porque los procesos de búsqueda de empleo son muy prolongados (alta duración) aunque la incidencia sea baja, o puede deberse a que hay una alta frecuencia de desempleo (alta incidencia) aunque con baja duración.

Las estimaciones anteriores nos permiten hacer inferencias sobre las tasas de incidencia del desempleo por grupos de la población y de esa forma contribuir a la caracterización del desempleo de los mismos grupos. La incidencia se puede estimar como la razón de la tasa de desempleo dividida por la duración media de desempleo. Dichas estimaciones se presentan en el cuadro 12 a continuación.

Las estimaciones indican que la incidencia disminuye con la edad y con los niveles de educación. Es decir, a medida que aumenta la edad de las personas, la frecuencia con la que quedan desempleados disminuye. Lo mismo ocurre a medida que aumentan los niveles de educación. A partir de la información presentada se puede caracterizar el desempleo de los diferentes grupos.

En general, la conclusión que se obtiene es que la duración del desempleo aumenta con la edad y con los niveles educativos, mientras que la incidencia disminuye. En otras palabras, el desempleo de los jóvenes es un desempleo de corta duración, pero las instancias de desempleo son muy frecuentes. Esto es consistente con situaciones en las que hay poca experiencia laboral y poco compromiso con el mercado laboral.

Por el contrario, el desempleo de las personas mayores es de baja incidencia, pero de larga duración. Esto puede ser consistente con situaciones de obsolescencia de las competencias (por ejemplo por cambios tecnológicos), con situaciones de discriminación y con otras que hacen difícil para las personas obtener ofertas de empleo razonables (altos salarios de reserva).

Cuadro 12

Tasas de incidencia del desempleo									
		Hombres				Mujeres			
Años	Años	No Jefes de Hogar		Jefes de Hogar		No Jefes de Hogar		Jefes de Hogar	
Edad	Educ	Solteros	Casados	Solteros	Casados	Solteras	Casadas	Solteras	Casadas
25	0	10,36%	12,47%	31,83%	13,81%	5,00%	5,04%	5,02%	4,75%
	5	7,55%	7,97%	6,94%	8,25%	3,87%	3,57%	3,47%	3,65%
	11	3,75%	3,82%	3,80%	4,17%	2,70%	2,53%	2,53%	2,62%
	16	1,70%	1,74%	1,76%	1,96%	1,99%	1,89%	1,92%	2,01%
35	0	9,55%	10,50%	8,94%	10,71%	3,71%	3,39%	3,33%	3,37%
	5	6,73%	6,86%	6,41%	6,93%	2,71%	2,86%	2,52%	2,57%
	11	3,36%	3,37%	3,36%	3,52%	1,92%	3,49%	1,79%	1,86%
	16	1,45%	1,43%	1,53%	1,43%	1,44%	2,68%	1,38%	1,41%
45	0	6,65%	6,84%	5,72%	6,96%	2,44%	2,35%	2,42%	2,35%
	5	3,91%	4,37%	3,90%	4,58%	1,87%	1,73%	1,80%	1,79%
	11	2,22%	2,20%	2,15%	2,26%	1,34%	1,23%	1,30%	1,29%
	16	1,02%	1,03%	1,00%	0,96%	0,99%	0,93%	0,96%	0,99%
55	0	3,51%	3,58%	3,10%	3,42%	1,73%	1,62%	1,63%	1,64%
	5	2,23%	2,26%	2,28%	2,23%	1,30%	1,19%	1,23%	1,23%
	11	1,17%	1,12%	1,11%	1,17%	0,93%	0,85%	0,89%	0,89%
	16	0,56%	0,45%	0,58%	0,41%	0,68%	0,64%	0,67%	0,65%
Promedios Simples por Años de Edad									
25		5,84%	6,50%	11,08%	7,05%	3,39%	3,26%	3,23%	3,26%
35		5,27%	5,54%	5,06%	5,65%	2,44%	3,10%	2,26%	2,30%
45		3,45%	3,61%	3,19%	3,69%	1,66%	1,56%	1,62%	1,60%
55		1,87%	1,85%	1,77%	1,81%	1,16%	1,08%	1,10%	1,10%
Promedios Simples por Años de Educación									
	0	7,52%	8,35%	12,40%	8,72%	3,22%	3,10%	3,10%	3,02%
	5	5,11%	5,36%	4,88%	5,49%	2,44%	2,34%	2,25%	2,31%
	11	2,62%	2,63%	2,60%	2,78%	1,72%	2,02%	1,63%	1,66%
	16	1,18%	1,16%	1,22%	1,19%	1,28%	1,53%	1,23%	1,27%

Los gráficos siguientes, contruidos a partir de la información presentada, ilustran estos puntos. Como se puede ver, la tasa de incidencia disminuye a medida que aumenta la edad, mientras que la duración del desempleo aumenta. En el caso de los hombres jefes de hogar, la incidencia de los de 55 años de edad es un tercio de la de los que tienen 25 años, pero la duración es tres veces mayor

Gráfico 7

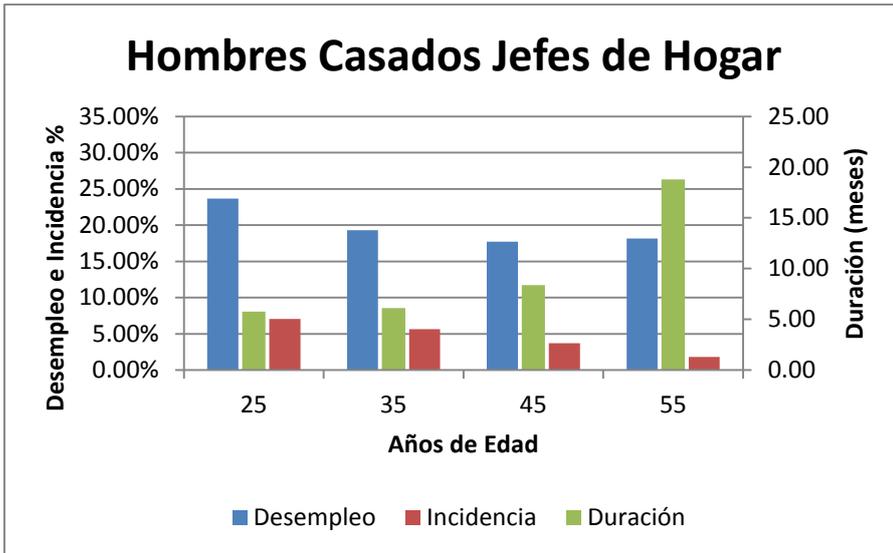
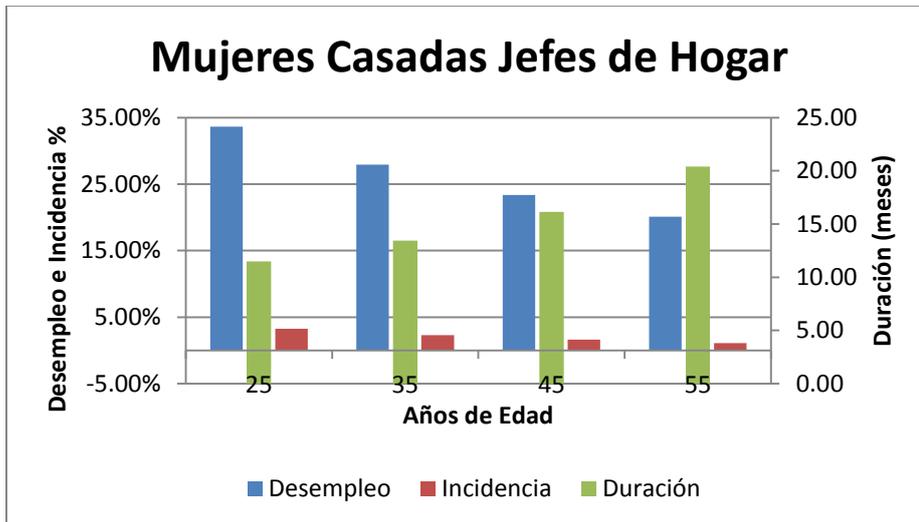


Gráfico 8



Algo similar ocurre por niveles educativos. Los trabajadores con bajos niveles quedan desempleados con mucha frecuencia pero duran desempleados poco tiempo. Los de altos niveles educativos quedan desempleados con baja frecuencia pero sus procesos de búsqueda son mucho más largos.

Lo interesante del resultado obtenido es que las políticas necesarias para combatir el desempleo de corta duración y de larga duración son diferentes. Generalmente, el desempleo de larga duración requiere procesos de retreinamiento y, en el caso de los trabajadores altamente calificados, de asistencia con información especializada sobre los nichos del mercado laboral relevantes. El desempleo de corta duración muchas veces puede ser disminuido con políticas activas de generación de empleo.

BIBLIOGRAFÍA

- **Allison, P. D.** (2004) Survival Analysis Using SAS, A Practical Guide, SAS Publishing, 7th Edition.
- **Baltagi, B. H.** (2001), A Companion to Theoretical Econometrics, Blackwell Publishers.
- **Cameron, A. and P. Trivedi** (2005), Microeconometrics: Methods and Applications, Cambridge University Press.
- **Cleves, M.A.** (2008), An Introduction to Survival Analysis Using Stata, A Stata Press Publication.
- **Castellar, C y Uribe, J. I. (2003)** “Determinantes de la duración del desempleo en el Área Metropolitana de Cali, 1988-2000”, *Archivo de Economía*, No. 218. Departamento Nacional de Planeación.
- **Ehrenberg, R. y Oaxaca, R. (1976)** “Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain” *AER*
- **Evans, M., N. Hastings, B. Peacock and C. Forbes (2011)**, Statistical Distributions, Fourth Edition, Wiley.
- **Fitzenberger, B. and Wilke, Ralf A. (2005)** “Using Quantile Regression for Duration Analysis”. *ZEW Discussion Paper,05-58, Mannheim*
- **Heckman, J. and B. Singer** (1984), “A method for minimizing the impact of distributional assumptions in econometric models for duration data”, *Econometrica*, 52.
- **Heij, C., de Boer, P., Franses, P. H., t. Kloek and H. van Dijk** (2004), Econometric Methods with Applications in Business and Economics, Oxford university Press.
- **Hosmer, D., Lemeshow, S. and May, S.** (2008) Applied Survival Analysis, Willey Series in Probability and Statistics, 2th Edition.
- **Klein, J. P. and Moeschberger, M. L.** (2003) Survival for Censored and Truncated Data, Springer, 2th Edition.
- **Lancaster, T. (1979)** “Econometric Methods for the Duration of Unemployment” *Econometrica*, Vol.47, No.4, pp. 939-956
- **Maddala, G.S. (1997)**, Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press.
- **Martínez, H. (2003)** “Cuánto duran los colombianos en el desempleo y en el empleo: un análisis de supervivencia”, *Archivos de Economía* 236
- **Meyer, B. (1990)** “Unemployment Insurance and Unemployment Spells” *Econometrica*, Vol.58, No.4, pp. 757-782

- **Núñez J. y R. Bernal** (1998), “El desempleo en Colombia: tasa natural, desempleo cíclico y estructural y la duración del desempleo (1976-1998)”, DNP, Archivos de Macroeconomía, Documento No. 97, Bogotá.
- **Pesaran, M. H. and P. Schmidt** (1999), Handbook of Applied Econometrics, Volume II: Microeconometrics, Blackwell Handbooks in Economics.
- **Pollmann-Schult, M. and Büchel, F. (2005)** “Unemployment Benefits, Unemployment Duration and Subsequent Job Quality: Evidence from West Germany” *Acta Sociologica* Vol. 48 No. 1, pp. 21-39
- **Tenjo, J. y Ribero, R. (1998)** “Participación, desempleo y mercados laborales en Colombia”, DNP, *Archivos de Macroeconomía, Documento No. 81*, Bogotá
- **van den Berg, G., Lindeboom, M and Dolton, P. (2006)** “Survey Non-Response and the Duration of Unemployment” *Royal Statistical Society*
- **Viáfara C. y Uribe, I. (2008)** "Duración del Desempleo y Canales de Búsqueda de Empleo en Colombia 2006" *Revista de Economía Institucional. Vol 11, # 21, 2009. Universidad Externado de Colombia.*
- **Wichert, L. and Wilke, R. (2008)** “Simple non-parametric estimators for unemployment duration analysis” *Royal Statistical Society Vol. 57, Part1, pp117-126*
- **Wooldridge, J. M.** (2001), Econometric Analysis of Cross Section and Panel Data, The MIT Press

Anexo 1

Supongamos que la variable aleatoria T que mide el tiempo de sobrevivencia o duración de un evento como, en el presente estudio, el desempleo. La probabilidad de que un sujeto seleccionado de manera aleatoria tenga un tiempo de sobrevivencia T menor que un tiempo específico t se define como $F(t) = P(T \leq t)$. De tal forma que, la función de sobrevivencia está asociada a la probabilidad de observar un tiempo de sobrevivencia (desempleo en este caso) mayor que dicho t , $S(t) = P(T > t)$, es decir, $S(t) = 1 - F(t)$

En el contexto de modelos de sobrevivencia, el objetivo econométrico es determinar qué tan largo es el período de vida de los sujetos bajo estudio. Es decir, en nuestro caso, determinar que tan larga es la duración del desempleo. Con este propósito, se utiliza el estimador no paramétrico de la función de sobrevivencia de Kaplan-Meier. La construcción de esta estadística se basa en una muestra conformada por n duplas (t_i, c_i) tomadas de forma aleatoria de la distribución subyacente de la variable de sobrevivencia T y de una variable aleatoria de censura C . Se define $m \leq n$ tiempos de fracaso o muerte (o cambio de estado) y $n - m$ valores censurados o valores que permanecen en el estado bajo estudio. Con base en esta información se construye, sobre los tiempos de duración, la siguiente estadística de orden: $t_{(1)}, t_{(2)}, \dots, t_{(n)}$. Este conjunto de información posibilita la construcción de la estadística de Kaplan-Meier, ecuación (10).

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} \quad (10)$$

Donde: n_i corresponde al número de individuos en riesgo de muerte o cambio de estado al momento $t_{(i)}$ y d_i el número de muertes o de cambios de estado observados al momento $t_{(i)}$. La variable de censura c_i caracteriza a cada uno de los individuos de la muestra. En el caso en que $t < t_{(1)}$ entonces se define $\hat{S}(t) = 1$

El estimador de la varianza para la estadística de Kaplan-Meier utilizado en este trabajo es el estimador de varianza de Greenwood, el cual se define mediante la ecuación (11).

$$\hat{Var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (11)$$

El estimador del $100p$ -ésimo cuartil se define como el mínimo t_j para el cual $\hat{S}(t_j) < 1 - p$. Es decir, $q_p = \min\{t_j | \hat{S}(t_j) < 1 - p\}$

En los resultados del trabajo se presenta la media como medida de tendencia central y su desviación estándar. Sin embargo, es de señalar que en el contexto de sobrevivencia dicha

estadística presenta sesgo y por tanto, es más adecuado interpretar las estadísticas cuantílicas²⁰. Las ecuaciones (12) y (13) presentan la media y su varianza, respectivamente.

$$\hat{\mu}(t_{(m)}) = \sum_{i=1}^m \hat{S}(t_{(i-1)}) (t_{(i)} - t_{(i-1)}) \quad (12)$$

Donde $\hat{S}(t_{(0)}) = 1$ y $t_{(0)} = 0$

$$\hat{Var}(\hat{\mu}(t_{(m)})) = \frac{n_d}{n_d - 1} \sum_{i=1}^{m-1} \frac{A_i^2 d_i}{n_i (n_i - d_i)} \quad (13)$$

con

$$A_i = \sum_{j=i}^{m-1} \hat{S}(t_{(j)}) (t_{(j+1)} - t_{(j)})$$

$$n_d = \sum_{i=1}^m n_i$$

²⁰ Véase, Cameron y Trivedi (2005).

Anexo 2

Glosario de Variables

Variable dependiente: duración

censura: 0 si está desempleado
1 si está empleado

Covariables:

sm : Género
1 Hombre
0 Mujer
Edad : años de edad
Edad2 : edad al cuadrado
Estcivil : Estado civil
1 casado
0 No casado
Jhogar : Jefe de hogar
1 si es jefe de hogar
0 no es jefe de hogar
aeduc : Años de educación aprobados
aeduc2 : Años de educación aprobados al cuadrado
Fnmenor : Número de menores de 7 años a cargo
Fnmayor_1 : Número de personas entre 7 y 17 años estudiantes
Fnmayor_2 : Número de personas entre 7 y 22 años estudiantes { Excluyentes
Fmayor_acargo (o maycarga) : Número de personas mayores de 60 años que no reciben ingreso alguno
Lingf_tot_si = Lningreso : Logaritmo del ingreso familiar total descontando ingresos del individuo + ingreso del resto de la familia
Lingf_tot_si2 : variable Lingf_tot_si al cuadrado
Interacciones:
estcivil_jhogar=ecivxjhog: interacción entre estado civil y jefe de hogar
Lingf_tot_jhogar=ingfxjhog: interacción entre el ingreso del resto de la familia y jefe de hogar
Lingf_tot_fmayor_acargo=ingxmcarga : interacción entre ingreso del resto de la familia y personas mayores de 62 años a cargo
ingxfnmayor : interacción entre ingreso del resto de la familia y personas estudiantes entre 7 y 17 años de edad.