

Duración del Desempleo

Informe Final

Septiembre 2009

Felipe Lozano

felipe.lozano@lumnifinance.com

Resumen

El presente artículo pretende hacer una revisión de las principales variables que afectan la decisión de entrada en el mercado laboral y en la duración del estado de desempleo usando datos de CASEN 2006 y comparándolos con los datos de CASEN 2000. Se hace un análisis de supervivencia sobre la variable de cuánto tiempo lleva buscando empleo de dicha encuesta.

Se encuentra que variables descriptivas del hogar: parentesco con el jefe, número de personas, educación del jefe y nivel de ingresos, afectan la duración del desempleo. Además se encuentra que la duración del desempleo es mayor a medida que aumenta el nivel educativo. Sin embargo, durante la crisis de 1999, reflejada en los datos de CASEN 2000, la relación con la mayor parte de las características sociodemográficas no fue significativa. También se encontró que no existe una diferencia en la duración del desempleo para los universitarios, lo que implicaría que la crisis no los afectó en igual medida o que se adaptaron más rápido que el resto de la población.

Tabla de contenido

Introducción y Revisión de Literatura	3
Hechos estilizados sobre el desempleo y la participación	4
El Desempleo, la Participación y el Crecimiento	4
Explicando la Participación	5
¿Cuántas semanas buscó o ha estado buscando trabajo?	8
Tiempo Buscando Trabajo por Género y Nivel de Estudios	9
Otras Variables	12
Cox PHM	13
Metodología	13
Resultados Cox PHM-2006	14
Resultados Cox PHM-2000	16
Estimaciones Paramétricas.....	18
Selección Modelo Paramétrico.....	19
Estimación de la Duración del Desempleo	21
Conclusiones.....	22
Referencias	23
ANEXOS.....	24
A.1 Gráficas de Supervivencia para otras variables	24
A.2 Construcción del Modelo Duración Desempleo.....	25
A.3 Estimaciones Paramétricas	29
Regresión Log Logística.....	29

Introducción y Revisión de Literatura

Hacer un seguimiento de la variable de duración del desempleo requiere del hacer uso de técnicas econométricas como el Análisis de Supervivencia (AS). En el AS existe un conjunto de individuos que presentan un estado –en este caso estar desempleado– y se registra el tiempo que toma al individuos cambiar de estado. El principal objetivo es describir la función acumulativa de probabilidad, que determine la salida del estado inicial después de un determinado tiempo, condicionado a ciertas co-variables (*covariates*). Cuando los individuos no alcanzan a cambiar de estado durante el periodo de análisis se habla de observaciones censuradas. En el caso de las muestras de corte transversal, como CASEN, se cuenta con el tiempo hasta el momento de la encuesta en el que el individuo a estado desempleado. Aunque los problemas relacionados con este inconveniente han sido descritos (van Es, et Al 2000), el presente artículo asume que los tiempos en los que una persona lleva buscando son los tiempos de cambio del estado.

Para Chile existen varios estudios que tratan de describir el desempleo y sus determinantes. Cowan, et Al. (2003) al igual que Gatica J. y Romaguera, P. (2005), tratan de encontrar cambios estructurales que puedan estar afectando el nivel de desempleo después de la crisis de 1998. Encuentran que niveles de desempleo sobre 10% no son un fenómeno nuevo en Chile y que ante shocks reales negativos los salarios caen comparativamente poco, con baja respuesta de los salarios reales. No encuentran evidencia concreta de cambios estructurales, pero sí del comportamiento cíclico de la oferta laboral y del desempleo. Cowan, et Al (2003) describiendo el comportamiento del desempleo, encuentran que el aumento en el desempleo se debe a la caída en la creación de empleo no acompañada por una caída en el crecimiento de la población económicamente activa. La población más afectada es la de los jóvenes (18-25 años) con poca experiencia y con educación relativamente alta (media, CFT o IP).

Además de lo anterior Gatica y Romaguera (2005), discuten nuevos temas de economía laboral: la rotación del empleo, las reformas laborales y el proceso de inserción en el mercado laboral de grupos específicos, como los jóvenes. Encuentran que durante la crisis hubo un deterioro en la calidad del empleo con una migración a formas contractuales que ofrecían menor protección a los trabajadores, un aumento de la informalidad y de los trabajadores por cuenta propia. Haciendo un análisis simple de supervivencia encuentran que después de un año la probabilidad de mantenerse en un contrato laboral está por debajo del 50%, pero en algunos sectores no supera el 30%. Concluyen que la rigidez del salario es causa de la destrucción de empleo y del ajuste por cantidad en tiempos de crisis. Sin embargo, no explican el origen de la rigidez de los salarios de forma concluyente y sugieren mayores investigaciones al respecto.

El estudio de Fuenzalida y Ruiz-Tagle (2008) trata de analizar el riesgo de moratoria de los hogares chilenos. El estado de moratoria depende del nivel de endeudamiento del hogar al igual que de la estabilidad de los ingresos de los hogares, que en su mayoría provienen del mercado laboral. Es en ese contexto que se analiza la duración del empleo bajo la probabilidad de quedar desempleado. Encuentran que los hombres

tienen una mayor probabilidad de permanecer empleados que las mujeres, independiente de la edad y del nivel de estudios. El nivel educativo también afecta de forma significativa con una probabilidad 45% menor de perder el empleo para la población con ES en comparación con los estudiantes que sólo tienen primaria y 27% en comparación con los que tienen educación media.

El análisis gráfico de la duración del desempleo es conducido con curvas de supervivencia producidas con el estimador empírico de Kaplan-Meier. Algunas pruebas de bondad de ajuste son realizadas para asesorar la distribución de las funciones de supervivencia. Se comparan los resultados para el año 2000 y para el 2006 para contrastar la duración del desempleo durante dos momentos del ciclo económico: desaceleración (2000) y expansión (2006).

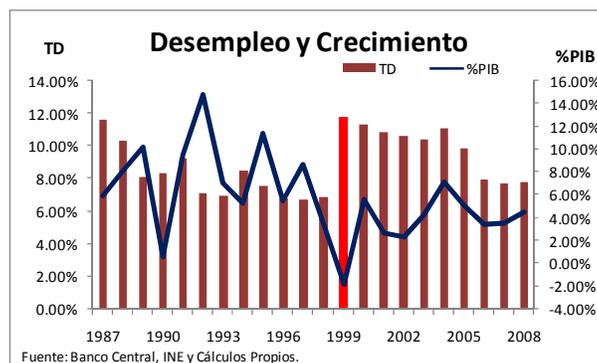
En la siguiente sección se analizan gráficamente las variables disponibles en CASEN 06 sobre la participación en el mercado laboral y la duración del desempleo. Posteriormente, se presentan las estimaciones del Análisis de Supervivencia para el desempleo siguiendo la forma no paramétrica convencional, el modelo *Cox Proportional Hazards*. A partir de esas estimaciones se construyó un modelo de variables independientes, pero debido a que no se pudieron rechazar las hipótesis de proporcionalidad se adoptaron soluciones paramétricas. Las estimaciones paramétricas son el tema de la sección 5, de la selección de la forma funcional y de las estimaciones de la duración del desempleo final. La sección 6 concluye.

Hechos estilizados sobre el desempleo y la participación

Antes de analizar la base de CASEN 06 nos detenemos a analizar el comportamiento histórico de las series de Participación y de la Tasa de Desempleo en Chile durante los últimos años con datos del Banco Central de Chile y del INE.

El Desempleo, la Participación y el Crecimiento

Figura 1. Desempleo y Crecimiento (1987 – 2008)

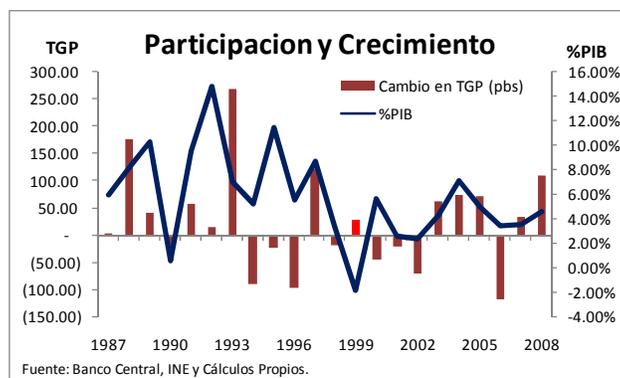


La Figura 1 ilustra la relación entre el desempleo y el ciclo económico en Chile, particularmente, el efecto de la crisis de 1999. Las series de crecimiento del PIB corresponden al crecimiento anual terminado en el tercer trimestre y las de desempleo corresponden al nivel promedio del tercer trimestre. La economía chilena entró a la década desacelerada, producto de la crisis económica mundial de fin de la década pasada y a lo largo de la década luchó por recuperar los niveles de desempleo y de crecimiento de los 90. Mientras que el promedio de crecimiento 91-99 fue de 7.11%, el correspondiente al periodo 00-08 fue de 4.3%. De la misma forma el desempleo promedio fue 7.85% y 9.74%, para los 90 y lo corrido de esta década, respectivamente.

Un estudio de la duración del desempleo debe tener en cuenta el papel de las crisis económicas en la duración del desempleo. En el presente artículo se comparan los resultados para los datos de las encuestas de CASEN 2000 y CASEN 2006 con ese propósito. La primera encuesta se hace en medio de una crisis aguda y la segunda en un año de crecimiento promedio.

Cowan, et Al. (2003) encuentran que la participación en el mercado laboral tiene un comportamiento cíclico y que tiende a disminuir en los períodos de recesión. Lo anterior lo explican con una correspondencia en el aumento en la escolaridad de los jóvenes que salen del mercado laboral en recesiones. Como se puede apreciar en la Figura 2, las desaceleraciones (pendiente negativa) son correspondidas en la mayor parte de los casos con disminuciones en la tasa global de participación; al igual que las aceleraciones con aumentos en la TGP. De todas formas no es claro el efecto sobre la participación que se encontró en niveles muy cercanos en el promedio para las dos décadas: 54.1% y 54.4%, para la 91-99 y 00-08, respectivamente.

Figura 2. Participación y Crecimiento (1987 – 2008)

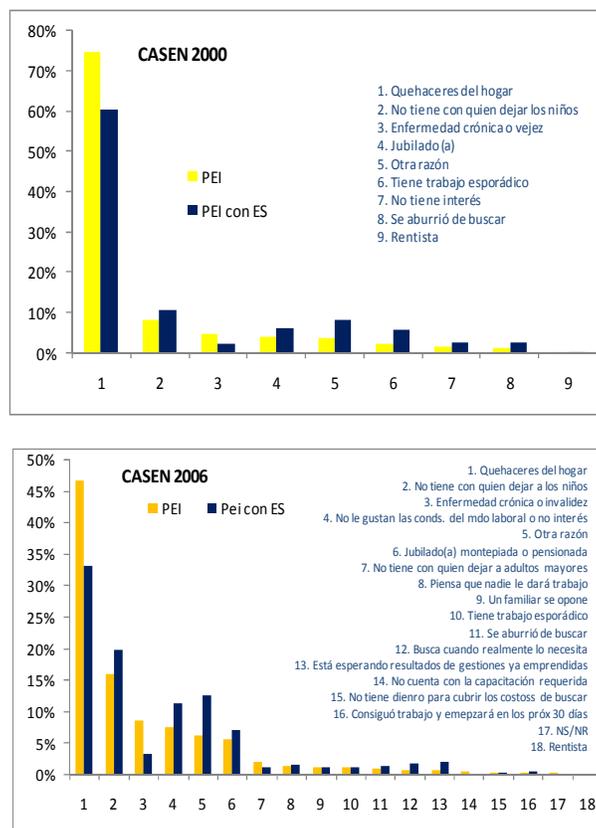


Explicando la Participación

En la presente sección se analiza la pregunta de por qué una persona inactiva, no ha buscado trabajo durante las últimas cuatro semanas. En el 2006 la mayoría de la población económicamente inactiva se encuentra estudiando: para toda la población el porcentaje es del 42% mientras que en el grupo de inactivos con o en ES, es del 68%. Los estudiantes fueron excluidos del siguiente análisis. En el caso de CASEN 2000, esos porcentajes correspondieron al 38% y al 65%.

En el caso del género, con una participación del 73% para hombres y de 39% para las mujeres –en el 2000 la participación masculina fue del 72% y la femenina del 33%–, las razones atribuidas a la no participación deben ser analizadas por género independientemente. La Figura 3 y la Figura 4 presentan los motivos para no buscar trabajo, sin incluir a los que se encontraban estudiando para toda la población y para la población con educación superior. Las categorías fueron diferentes en los años analizados. Los datos están ordenados por la participación de cada una de las causas en el caso de la población económicamente inactiva (PEI) total.

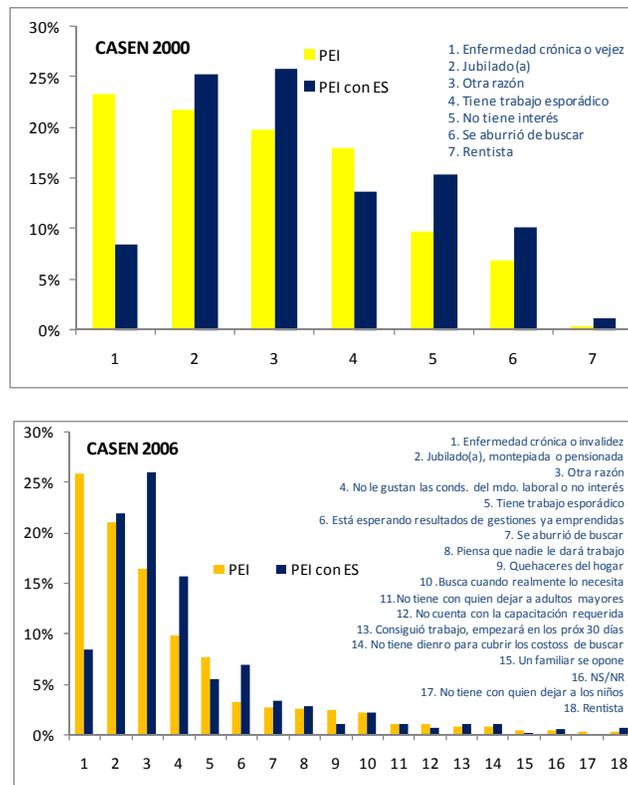
Figura 3 ¿Por qué no buscan trabajo? – Mujeres



Viendo las principales causas de no participación en el caso de las mujeres, los quehaceres del hogar y el cuidado de los niños retienen al 63% de participar en el mercado laboral, aunque esta proporción cae hasta el 52% en el caso de mujeres con ES. Estos porcentajes han caído drásticamente pues en 2000 correspondían al 75% de todas las mujeres inactivas y al 61% en el caso de las mujeres con ES. Estas causas se encuentran no son tan relevantes en el caso de los hombres: 3% para toda la población y apenas 1% para aquellos con ES y no hay reportes en el 2000. Motivos de salud e incapacidad son la siguiente razón para las mujeres y la primera causa de inactividad entre los hombres, con 9% y 26%, respectivamente para el 2006, en el caso de toda la PEI –en el 2000, 5% y 23%–. En el caso de la PEI con ES los motivos de salud representaron el 3% y el 8%, para mujeres y hombres –en el 2000, 8% y 2%–. Es de resaltar la falta de peso de esta razón en la población con ES, indicando

algún tipo de relación entre el nivel de educación y el riesgo de sufrir enfermedades que incapaciten a la población en edad de trabajar.

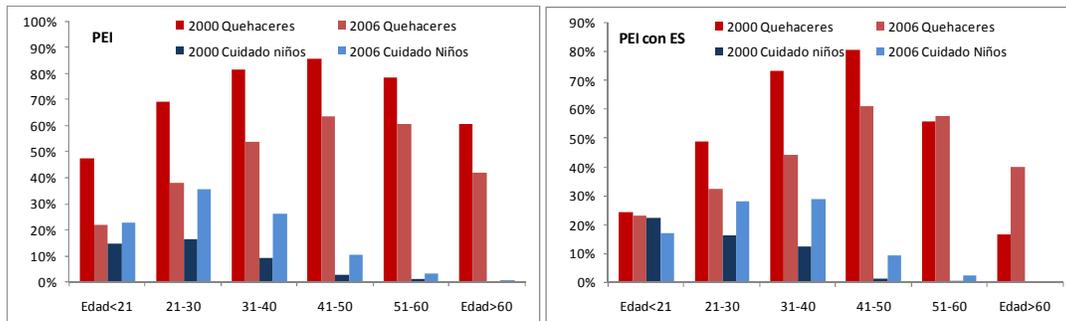
Figura 4 Por qué no buscan trabajo – Hombres



En el caso de la población con ES los motivos tienen diferentes pesos a los de la PEI. En el caso de los hombres no queda claro cuál es el principal motivo, dado que el 26% atribuyen la inactividad en la búsqueda de empleo a “Otra Razón”, seguido por los jubilados antes de los 65 años que ocupan el 22% de los hombres –en el 2000, 25% tienen algún tipo de pensión y el 26% tiene otra razón para no participar–. Las mujeres con ES también tienen esa diferencia después de tener en cuenta el cuidado del hogar y de los hijos.

La evolución de las principales causas que alejan a las mujeres del mercado laboral, los quehaceres domésticos y el cuidado de los niños, se presenta la Figura 5. En el lapso de 6 años entre las dos encuestas ha habido una reducción en la importancia de estas dos causas. La reducción ha sido mayor en el caso de la PEI porque empezó con una base comparable más elevada. El aumento en la proporción de mujeres que no participan por el cuidado de los niños es relativo, al disminuir la proporción de mujeres que por quehaceres no participaban, todas las demás razones aumentaron su capacidad de explicar la inactividad económica. En los grupos por edades, igualmente, se puede apreciar que las mujeres con ES posponen la decisión de ser madres hasta más allá de los 30 años en mayor proporción que el resto de la población.

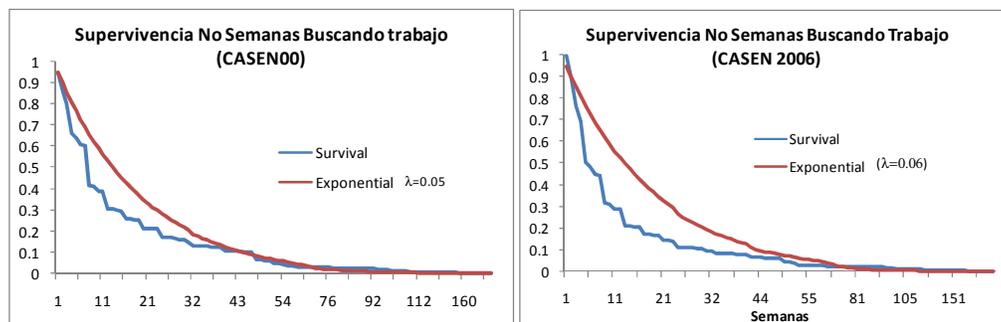
Figura 5. Evolución principales causas de inactividad en las mujeres por grupos de edad



¿Cuántas semanas buscó o ha estado buscando trabajo?

En esta sección se presenta el análisis gráfico de las curvas de supervivencia (Kaplan-Meier)¹ de las bases analizadas y con respecto a las variables categóricas de género y de nivel de estudios. Las gráficas de otras variables categóricas se encuentran en el anexo como pruebas de paralelismo para luego estimar las funciones de riesgo. La situación laboral de una persona es algo complejo, con un resultado no necesariamente binario. La salida de la situación del desempleo, puede darse tanto porque un individuo encuentra trabajo como por el hecho de que la persona se aburríó de buscar y prefirió dedicarse a otras actividades como los quehaceres del hogar o el cuidado de su familia. Por restricciones de información, en este artículo nos limitamos a analizar el tiempo promedio que una persona busca trabajo.

Figura 6. Funciones de Supervivencia



En la Figura 6 se puede apreciar la función de supervivencia para las diferentes versiones de CASEN. Los datos superiores a 208 semanas (4 años) fueron agrupados en un mismo intervalo para evitar el efecto de observaciones extremas. La gráfica correspondiente al año 2000 tiene una cola con mayor probabilidad, implicando que el promedio del tiempo que las personas estaban buscando trabajo era mayor que en 2006. En

¹ En las funciones de supervivencia empírica Kaplan–Meier, $S(t) = 1 - F(t) = \prod_1^t \frac{n_i - d_i}{n_i}$, donde $S(t)$ es la función de supervivencia; $F(t)$, es la función acumulativa de probabilidad; n_i , es el número de observaciones en riesgo en i , y d_i , es el número de observaciones que cambiaron de estado en el intervalo de tiempo i .

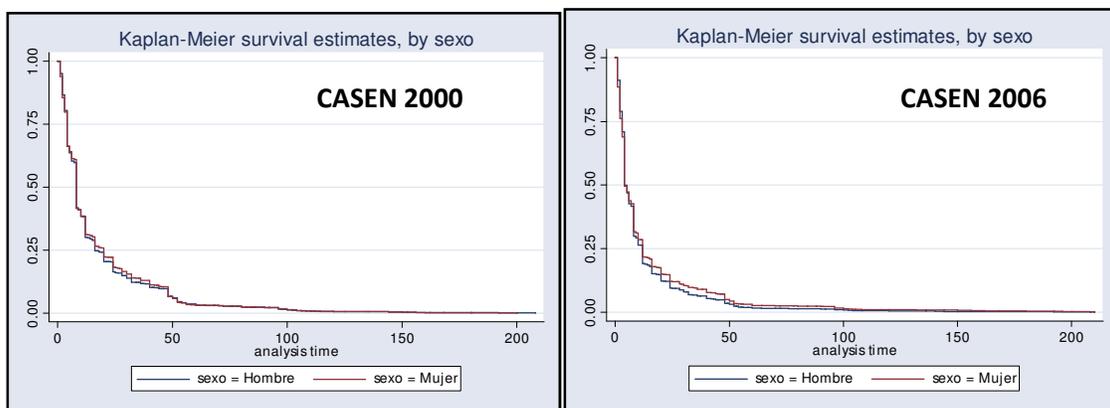
el año 2000, la probabilidad de que una persona buscara trabajo por más de 24 semanas correspondía al 20%; para el 2006, esa misma probabilidad correspondía apenas a 15 semanas. En el 2000, el 50% de las personas habían buscado trabajo hasta 9 semanas, mientras que en 2006 apenas hasta 5 semanas. Con este resultado las estimaciones sobre la base CASEN 2000, son un estimado para tiempos de crisis y las de 2006 son un estimado conservador para años de expansión o de ausencia de crisis.

Se estimaron parámetros para distribuciones Poisson y Exponenciales. Se minimizó el valor la prueba de bondad de ajuste de Pearson para encontrar el parámetro óptimo de cada función. Los valores encontrados para la f.d. Poisson rechazaban la hipótesis nula de que la distribución de la función empírica de supervivencia tuviera esa forma. En las gráficas se reportan los resultados de la función de supervivencia de la distribución exponencial, y estos valores no permiten rechazar la hipótesis nula de esta forma de distribución, siempre y cuando se usen los intervalos empíricos y no todos los posibles valores de t entre 0 y 208. Debido a que la frecuencia de la duración del desempleo no es estrictamente monótona es probable que una función log normal o log logística se adapte mejor a los datos, ya que estas funciones permiten que, en función del tiempo, haya primero un crecimiento en las frecuencias estimadas y que luego caigan acorde. La estimación de dichas funciones se deja para integrarla a las variables explicativas en la sección de estimaciones paramétricas.

Tiempo Buscando Trabajo por Género y Nivel de Estudios

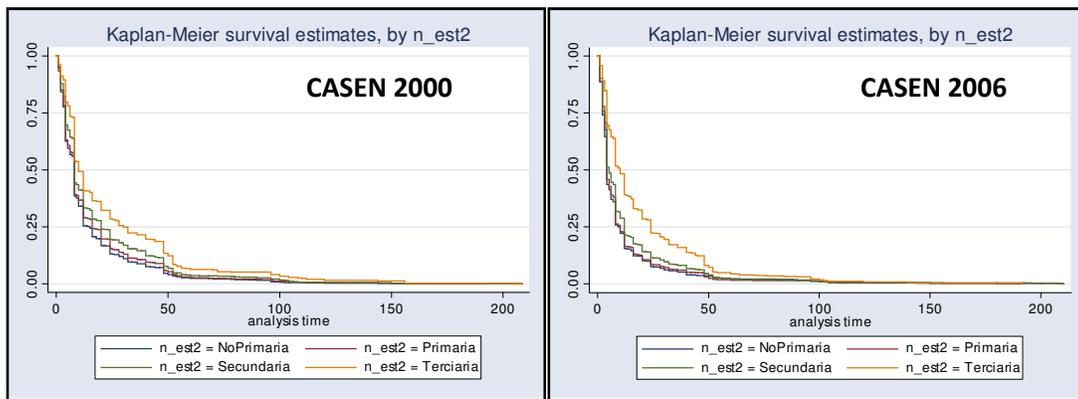
La Figura 7 presenta las estimaciones de la funciones Kaplan-Meier diferenciadas por género para toda la muestra. Las Figuras 8 y 9 presentan las gráficas por niveles de estudio, diferenciando entre primaria secundaria y terciaria, y luego sólo diferenciando por los niveles de la ES. En las funciones de supervivencia por género se puede apreciar que los hombres tienen una menor probabilidad de permanecer por largo tiempo desempleados, y que en 2006 la brecha es mayor que en el año 2000. Aunque nos habría podido interesar primero la duración del desempleo al primer trabajo, los resultados por tipo de educación no eran concluyentes ya que sólo se cuenta con 30 observaciones a nivel universitario, 24 para IP y 14 para CFT.

Figura 7. Funciones de Supervivencia por Género



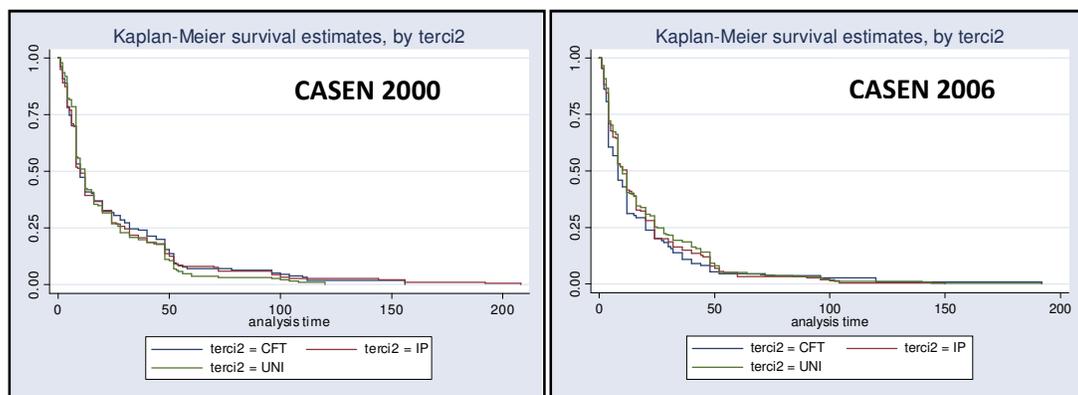
El nivel de estudio no parece garantizar que una persona deje de buscar trabajo más rápido. Tanto en el año 2000 como en el 2006 la población con ES tiene mayores tiempos de búsqueda de trabajo (Figura 8). El 80% de la muestra sin ningún nivel completo o con sólo primaria declaran hasta 13 semanas de búsqueda, la población con secundaria hasta 14 semanas y el grupo con educación terciaria hasta 28 semanas. Es probable que la población con ES siga buscando trabajo aun si no encuentra, mientras que el resto de la población abandona la búsqueda o está dispuesta a aceptar posiciones peor remuneradas más pronto.

Figura 8. Funciones de Supervivencia por Nivel de Estudios



Poniendo el foco sólo en la población con ES, parece que las circunstancias han cambiado. Al comienzo de la década, la población con estudios universitarios y superiores pasaba menor tiempo buscando trabajo, en promedio, que los graduados de IPs o CFTs, aunque la diferencia no parecía ser significativa. La relación se ha invertido en tiempos de búsqueda inferiores a un año, y no parece haber una diferencia significativa en periodos más largos de búsqueda.

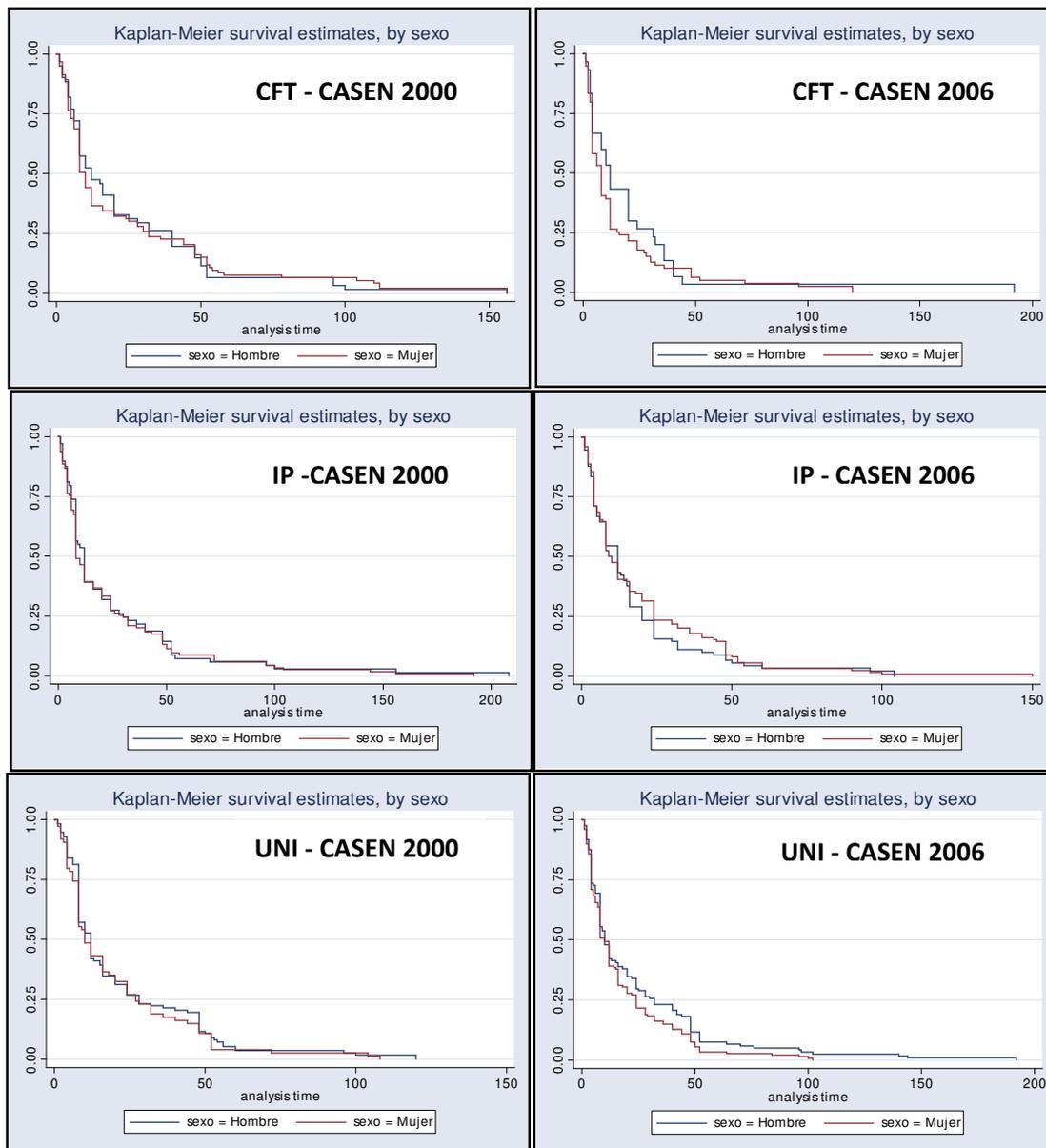
Figura 9. Funciones de Supervivencia para Niveles de ES



La colección de gráficas presentadas en la Figura 10 presentan las diferencias por género en los diferentes niveles de la educación superior. En el 2006, a nivel universitario las mujeres presentan menores

tiempos de búsqueda que los hombres; en el caso de los IPs sucede lo contrario y en el de los CFTs las mujeres tienen mayor probabilidad de buscar. En el año 2000 la distinguir entre género no parece tener un efecto significativo para los diferentes niveles de la educación superior, ya que en los tres niveles especificados las funciones de supervivencia de uno y otro género se encuentran entrelazadas.

Figura 10. Funciones de Supervivencia – Nivel de Estudios ES y Género



Otras Variables

Además de analizar gráficamente las variables categóricas en la sección anterior y en el Anexo A.1, se condujeron pruebas de ranking logarítmico para probar la igualdad entre las funciones de supervivencia. El *Logrank Test*² prueba la hipótesis nula de igualdad entre varias funciones estimadas de supervivencia. Si se rechaza la hipótesis nula, el efecto de la variable dicotómica es significativo y aporta al modelo final. Helmer y Lemeshow (1999) recomiendan un corte del *p-value* del 20%. La Tabla 1 presenta los resultados para el 2000 y la Tabla 2 para el 2006. La variable regiones –observaciones de individuos fuera de Santiago- resulta no aportar información al modelo en el 2000. Las demás variables, con excepción del nivel de estudios, también parecen perder poder explicativo en el 2000. Si se tiene en cuenta que en el 2000 se había incorporado completamente el ajuste post crisis asiática en el mercado laboral, y que en 2006 la economía estaba en una senda de crecimiento, los resultados parecen dar indicios de la diferencias en el mercado laboral para los dos momentos.

Tabla 1. Log Rank Test para Variables Categóricas – CASEN2000

Variable	df	chi2(df)	Pr>chi2
E. Civil	6.00	9.51	0.1469
Parent. Jefe	10.00	15.79	0.1058
Niv Est	3.00	109.60	0.0000
Regiones	1.00	1.09	0.2964
Genero	1.00	2.22	0.1363
Exper. Previa	1.00	0.19	0.6618

Tabla 2. Log Rank Test para Variables Categóricas – CASEN2006

Variable	df	chi2(df)	Pr>chi2
E. Civil	7.00	33.36	0.0000
Parent. Jefe	12.00	29.08	0.0038
Niv Est	3.00	115.14	0.0000
Regiones	1.00	17.17	0.0000
Genero	1.00	19.29	0.0000
Exper. Previa	1.00	0.08	0.7804

En el caso de las variables continuas se corrieron regresiones de Funciones de Riesgo Proporcionales (Cox Proportional Hazards Model - PHM) previas para cada una para evaluar su aporte al modelo. Los resultados se presentan en la Tabla 3 y anticipan el resultado de la sección de estimaciones donde se presenta el procedimiento de estimación. Sólo se presentan los coeficientes de las regresiones individuales y el *p-value* correspondiente. En el año 2000 la relación de parentesco, los años de educación, el ingreso no laboral en el hogar, la edad, y el nivel de estudios parecen aportar información al modelo. El nivel de ingresos del hogar o los ingresos laborales del individuo se descartan como variables que puedan influir en la explicación de la variable duración de búsqueda de empleo. Para el año 2006, el resultado sólo es no significativo para la variable del nivel de ingreso del hogar.

² El LogRank test se calcula a través de $LR = \mathbf{u}'\mathbf{V}^{-1}\mathbf{u}$. Donde el vector fila $\mathbf{u}' = \sum W(t_j)(d_{1j} - E1_{1j}, \dots, d_{rj} - E1_{rj})$; r es el número de categorías en la variable y la función $W(t_j) = 1$ si $n_{ij} > 0$, y $W(t_j) = 0$ si $n_{ij} = 0$. En el caso de la matriz $r \times r$ \mathbf{V} , cada elemento se obtiene a través de:

$$V_{ij} = \sum_{k=1}^k \frac{W^2(t_k)n_{ik}d_j(n_j - d_j)}{n_j(n_j - 1)} \left(\delta_{il} - \frac{n_{ij}}{n_j} \right); i=l, \dots, r; l=l, \dots, r; \text{ y } \delta_{il} = 1 \text{ si } i = l, \text{ de lo contrario } \delta_{il} = 0.$$
 La prueba usa información de todas las funciones de supervivencia construidas.

Tabla 3. Cox PHM - Variables Continuas Individuales

2000	Coef	P> z	2006	Coef	P> z
Ln (Ing Tot Hogar)	0.00	0.833	Ln (Ing Tot Hogar)	0.00	0.628
Ln (Ing NoLb Hogar)	-0.01	0.014	Ln (Ing NoLb Hogar)	-0.01	0.000
Ln (Ing NoLb Ind)	0.00	0.453	Ln (Ing NoLb Ind)	0.01	0.006
Educ Jefe	-0.01	0.000	Educ Jefe	-0.01	0.013
Ln Edad	-0.23	0.000	Ln Edad	-0.26	0.000
Numper	0.01	0.187	Numper	0.02	0.000
s	-0.02	0.000	s	-0.03	0.000

Cox PHM

Metodología

Las presentes estimaciones siguen la estimación propuesta por Cox PHM presentadas en el Capítulo 3 de Hosmer y Lemeshow (1999). La función de riesgo $h(t, \mathbf{x}, \boldsymbol{\beta})$ describe la probabilidad de seguir en el estado inicial después de haberse mantenido en él durante el tiempo t . En la estimación propuesta por Cox la función de riesgo se puede escribir como el producto de dos funciones independientes, en donde las co-variables son independientes del tiempo. En un caso lineal, la variable dependiente se podría modelar como $y = \alpha + \beta x_1 + \sigma \varepsilon^*$, donde $y = \ln(t)$ y $\varepsilon^* = \ln(\varepsilon)$. Donde la variable de tiempo en el estado de supervivencia puede ser escrita también como: $t = (e^{\alpha + \beta x_1}) \times \varepsilon^\sigma$.

Al separar la función que describe el tiempo de supervivencia en un componente sistemático, descrito por las variables independientes y en un componente de error permite hacer supuestos paramétricos sobre el componente de error y estimar separadamente los parámetros del componente sistemático. Al asumir formas paramétricas sobre el componente de error, como se hizo tangencialmente en la sección anterior con la función exponencial, se está asumiendo una estructura funcional para la función de riesgo.

El modelo semiparamétrico de Cox no hace un supuesto sobre la forma paramétrica de la función de riesgo pero mantiene el supuesto de que se pueden separar los dos componentes sobre la función. De tal forma que la función de riesgo se puede escribir como $h(t, \mathbf{x}, \boldsymbol{\beta}) = h_0(t)r(\mathbf{x}, \boldsymbol{\beta})$. La función $h_0(t)$ describe cómo cambia la función de riesgo en función del tiempo de supervivencia, y la función $r(\mathbf{x}, \boldsymbol{\beta})$ describe el cambio en función de las variables explicativas. De lo anterior se puede definir el radio de riesgo (*Hazard Ratio*):

$$HR(t, x_1, x_0) = \frac{h_0(t)r(x_1, \boldsymbol{\beta})}{h_0(t)r(x_0, \boldsymbol{\beta})} = \frac{r(x_1, \boldsymbol{\beta})}{r(x_0, \boldsymbol{\beta})} \quad (1)$$

La anterior definición permite hacer una estimación, después de hacer una suposición sobre la forma funcional de $r(\mathbf{x}, \boldsymbol{\beta})$ que es independiente de la forma funcional de la función de riesgo. Cox (1972) sugirió $r(\mathbf{x}, \boldsymbol{\beta}) = e^{x\boldsymbol{\beta}}$. Incorporando esa sugerencia el HR se puede definir el modelo a estimar Cox PHM como:

$$HR(t, x_1, x_0) = e^{\beta(x_1 - x_0)} \quad (2)$$

La estimación convencional para el modelo expresado en la Ecuación (2) es el de máxima verosimilitud. Al tener un estimado para todos los β de cada una de las variables independientes, la relación entre la función de riesgo y la función de supervivencia es $S(t, \mathbf{x}, \boldsymbol{\beta}) = e^{-r(\mathbf{x}, \boldsymbol{\beta})H_0(t)}$, donde la función cumulativa de riesgo $H(t, \mathbf{x}, \boldsymbol{\beta}) = r(\mathbf{x}, \boldsymbol{\beta}) \int_0^t h_0(u) du = r(\mathbf{x}, \boldsymbol{\beta})H_0(t)$. Si se describe la función de supervivencia base como $S_0(t) = e^{H_0(t)}$, y se reemplaza la parametrización sobre $r(\mathbf{x}, \boldsymbol{\beta})$, sugerida por Cox, entonces la función de supervivencia se puede escribir como:

$$S(t, \mathbf{x}, \boldsymbol{\beta}) = [S_0(t)]^{\exp(\mathbf{x}\boldsymbol{\beta})} \quad (3)$$

Una vez se ha establecido el proceso de estimación y la relación con las variables que hemos venido describiendo se presentan los resultados de la estimación en la siguiente sección.

Resultados Cox PHM-2006

En la Tabla 4 se presenta el modelo final para la regresión con base en CASEN 2006 y en el Anexo A.1, se presentan los pasos que se siguieron para la configuración del modelo final, incluyendo las pruebas de proporcionalidad para las variables incluidas en el análisis. Para analizar el efecto de las variables se presentan los resultados con sus *hazard ratios*.

Tabla 4. Cox PHM – CASEN 2006

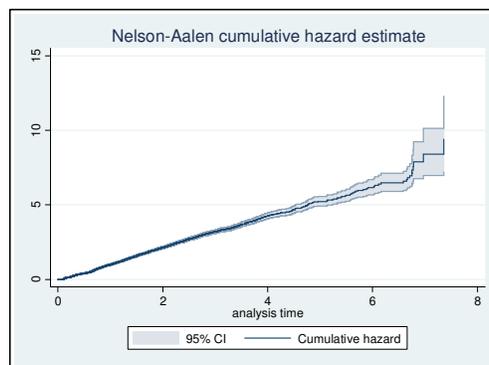
Number of obs	6,843					
LR chi2(13)	289.01					
Prob > chi2	0.00					
No. of failures	6,843					
Time at risk	79,794.00					
Log likelihood	-54,120.79					
	Haz. Ratio	Std. Err	z	P> z	[95% Conf. Int]	
Espos(a) o Pareja	1.1737	0.0437	4.30	0.000	1.091	1.263
Hijo(a)	0.8814	0.0283	-3.93	0.000	0.828	0.939
primaria	0.8866	0.0349	-3.06	0.002	0.821	0.958
secundaria	0.8648	0.0264	-4.76	0.000	0.815	0.918
CFT	0.8114	0.0829	-2.05	0.041	0.664	0.991
IP	0.7889	0.0611	-3.06	0.002	0.678	0.918
UNI	0.7143	0.0490	-4.90	0.000	0.624	0.817
sexo	0.9339	0.0250	-2.55	0.011	0.886	0.984
regiones	1.0653	0.0314	2.14	0.032	1.005	1.129
ln_ynl	1.0100	0.0028	3.56	0.000	1.004	1.016
ln_nolbh	0.9938	0.0039	-1.57	0.116	0.986	1.002
educ_jefe	1.0096	0.0037	2.60	0.009	1.002	1.017
ln_edad	0.6008	0.0247	-12.38	0.000	0.554	0.651
numper	1.0178	0.0070	2.56	0.010	1.004	1.032

Para las variables dicotómicas se encuentra que las parejas de los jefes de hogar salen del estado de desempleo un 17% más rápido que los jefes, mientras que los hijos que viven con sus padres se demoran en ese estado hasta un 12% más que los jefes de hogar. Las mujeres toman 6.7% más de tiempo que los hombres en el estado de desempleo y las personas fuera de la Región Metropolitana 6.5% menos que los santiaguinos.

Se analizaron variables dicotómicas para los diferentes niveles de educación completos y se descartó la variable de escolaridad medida en años por problemas de colinearidad. Esta forma permite aislar los efectos de los niveles completos sobre la duración del desempleo. Sin embargo, las variables definidas violaban el supuesto de proporcionalidad en mayor proporción que la variable de años de educación s. Considerando sólo los años de educación, cada año de educación por encima de la media muestral (10 años) aumenta el tiempo de búsqueda en 3.8%. Ya en el modelo final, se obtiene que una persona que ha terminado sólo el nivel de primaria sale del estado inicial de desempleo a una tasa 11.3% más lento que las personas sin ningún tipo de educación. Haber concluido cuarto medio implica una reducción extra de 13% sobre la tasa de salida del desempleo de los que terminaron primaria. En la educación superior los graduados universitarios son quienes tienen un estado de desempleo más largo con 0.71 de la tasa de la población educación secundaria, seguidos por los graduados de IPs (0.79) y CFTs (0.81). Este resultado está relacionado con el tiempo y los recursos que se invierten en la búsqueda de empleo por parte de la población con mayor educación; en cambio las condiciones precarias de las personas con menor educación pueden estar aceptando trabajos de menor remuneración y que impliquen menores costos para ellos en el proceso de búsqueda.

En el caso de la edad, medida en logaritmos, un cambio de un 100% sobre la media muestral (de 32 a 64 años) implica una caída en la probabilidad de salir del estado de desempleo del 40%. Tanto la edad como el nivel de educación parecen indicar que a cierto nivel de experiencia y/o educación la especialización laboral de los individuos es más elevada y al encontrarse en situación de desempleo, el rango de oficios que puedan hacer provecho de esa especialización es más limitado.

Figura 11. Nelson - Aalen Función Cumulativa de Riesgo (Residuos Cox Snell 2006)



El ingreso no laboral también tiene un efecto significativo sobre la duración del estado de desempleo. Un aumento del 100% en los ingresos no laborales hace que la tasa a la que se sale del estado inicial aumente en un 1%. Dado que la media muestral es CLP30/mes, ya que sólo un cuarto de las observaciones declara algún ingreso diferente de cero, y existiendo observaciones sobre CLP5millones, las observaciones con ingresos no laborales elevados, salen mucho más pronto del desempleo. Esta variable se podría estar comportando como una proxy de las relaciones sociales que existen para la población de mayores ingresos, que les permitiría salir

del estado de desempleo más rápido. Por otro lado, la variable de ingresos no laborales del hogar no presenta dicha polarización extrema y tiene un símbolo contrario, lo que refleja el efecto de la riqueza sobre las demás personas del hogar. Un aumento del 100% a partir de la media (CLP 50,000/mes) en los ingresos no laborales del hogar aumentan la el tiempo de duración en el desempleo en un 0.7%.

La Figura 11 muestra la función de riesgo acumulativa de Nelson Aalen para los errores Cox Snell de la estimación presentada en la Tabla 4, en escala logarítmica. Se puede apreciar que las estimaciones de las funciones de supervivencia generadas a partir del modelo de Cox son relativamente confiables dentro de un rango inferior a un año ($\ln(52) = 3.95$), posteriormente, los intervalos de confianza se amplían. Tres de las variables no rechazaron la hipótesis de que el coeficiente de interacción con el tiempo fuera no significativo: la variable dicotómica de la universidad, la educación del jefe y la variable de género. Al existir una relación significativa entre $r(x, \beta)$ y $H_0(t)$, no se pueden construir estimaciones a partir de un modelo que incluyera sus interacciones con los datos sin hacer supuestos en la forma funcional de la relación entre las dos funciones. Los resultados del test global de proporcionalidad se presentan en la Tabla 5, donde el número de variables que no rechazan la hipótesis de no proporcionalidad aumenta a 5. Ante la imposibilidad de resolver el problema de proporcionalidad, se plantearon soluciones paramétricas para la estimación final de la duración del desempleo. Pero se mantuvo la presentación de interrelación de las variables que no cambia sustancialmente en las demás estimaciones y que a partir de los coeficientes de COX se pueden hacer comparaciones en términos relativos a las medias de la muestra y con respecto a las otras variables.

Tabla 5 Test de Proporcionalidad 2006 – Modelo Final

	rho	chi2	df	Prob> chi2
Esposo(a) o Pareja	0.00	0.07	1.00	0.79
Hijo(a)	0.00	0.05	1.00	0.82
primaria	0.00	0.00	1.00	0.99
secundaria	0.03	4.82	1.00	0.03
CFT	0.02	2.27	1.00	0.13
IP	0.04	9.63	1.00	0.00
UNI	0.05	19.30	1.00	0.00
sexo	-0.03	7.12	1.00	0.01
regiones	0.00	0.00	1.00	0.99
ln_ynl	-0.01	0.52	1.00	0.47
ln_nolbh	0.00	0.08	1.00	0.77
educ_jefe	-0.03	5.67	1.00	0.02
ln_edad	-0.01	0.47	1.00	0.49
numper	0.00	0.05	1.00	0.82
Global Test		52.95	14.00	0.00

Resultados Cox PHM-2000

Los resultados de las estimaciones usando la base CASEN 2000 se presentan en la Tabla 6. El proceso de configuración del modelo se presenta en el Anexo A.2. Desde la construcción preliminar del modelo se habían descartado varias variables que habían sido significativas en el modelo para el 2006. Las diferentes categorías del estado civil, y la mayoría de las categorías de relación de parentesco con el jefe son no

significativas y excluidas del modelo final. Igual resultado se había encontrado para la variable de género. De la preselección se incluyó el ingreso no laboral del hogar, pero en el modelo en conjunto resultó no significativo. Lo anterior podría ser un indicio de que en periodos de crisis las relaciones sociales que caracterizan esas variables no tienen poder explicativo sobre la duración del desempleo, pero una prueba de esa hipótesis requeriría de datos de series de tiempo incluyendo varias crisis económicas.

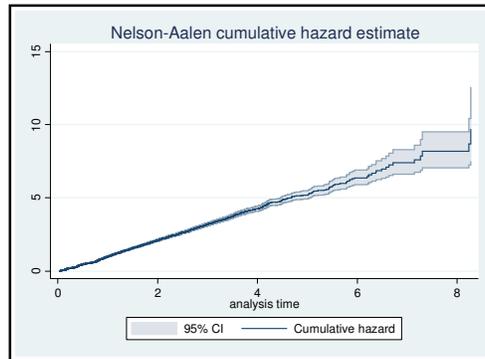
Las diferentes categorías educacionales arrojaron un resultado similar al encontrado en el 2006, y sólo dos niveles se mantuvieron no significativos: el nivel de primaria y el nivel universitario. El impacto sobre la duración del desempleo tiene el signo esperado pero el nivel es menor al encontrado en los otros niveles de la educación superior. Al descontar por su varianza el efecto es no significativo, indicando que en periodos de crisis la duración del desempleo es menor para el nivel universitario que para los otros niveles de educación, o incluso que no tiene un efecto en comparación con la población sin educación proclive a aceptar cualquier trabajo. Especialmente si es comparado con el anterior resultado para el 2006 donde se encontró que los graduados universitarios son quienes más tiempo se mantenían en buscando empleo. Para probar la hipótesis anterior se excluyó también la variable de años de educación y se corrió nuevamente el modelo. Se encuentra que el coeficiente que acompaña a la variable dicotómica para nivel universitario completo se vuelve significativo, pero sigue siendo menor que los otros niveles educativos con excepción del nivel de primaria.

Tabla 6 Cox PHM – CASEN 2000

No. of subjects	9,121					
LR chi2(8)	235.63					
Prob > chi2	0.00					
No. of failures	9,121					
Time at risk	144,815.00					
Log likelihood	-74,741.91					
	Haz. Ratio	Std. Err	z	P> z	[95% Conf. Int]	
primaria	0.9764	0.0426	-0.5500	0.5850	0.8963	1.0637
secundaria	0.8981	0.0325	-2.9700	0.0030	0.8366	0.9641
CFT	0.8347	0.0704	-2.1400	0.0320	0.7076	0.9847
IP	0.8617	0.0679	-1.8900	0.0590	0.7383	1.0057
UNI	0.9576	0.0804	-0.5200	0.6060	0.8123	1.1289
ln_ynl	1.0105	0.0055	1.9200	0.0550	0.9998	1.0213
ln_edad	0.6856	0.0221	-11.7000	0.0000	0.6436	0.7304
yrs_educ2	0.9810	0.0076	-2.4700	0.0140	0.9662	0.9961

Cada año educativo por encima de la media disminuye la probabilidad de salir del estado de desempleo en 2%, además el concluir cada nivel educativo disminuye esa probabilidad aún más, también porque los efectos de los niveles son cumulativos. Así una persona graduada de la educación superior también activa los coeficientes para primaria y secundaria anteriores. En el nivel de la educación Superior, según los coeficientes, el mayor tiempo registrado sería el de la población con educación técnica con un coeficiente de -18%, seguido por el de los graduados IP con un coeficiente igual a -15% y razón de riesgo (*hazard ratio*) de 0.861. Finalmente, la duración del desempleo del nivel universitario no parecería tener una diferencia significativa con la duración del desempleo de los graduados de la enseñanza media.

Figura 12 Nelson - Aalen Función Cumulativa de Riesgo (Residuos Cox Snell 2000)



El ingreso no laboral del individuo aumenta la probabilidad de salir del estado de desempleo en un 1% por cada vez que se doble el ingreso por encima de la media. La edad, en cambio, tiene un impacto negativo sobre el tiempo de búsqueda y al doblarse la edad la probabilidad de salir del estado de desempleo cae en un 37%. La Figura 12 muestra la función cumulativa de riesgo Nelson Aalen para el modelo estimado usando las estimaciones de residuos Cox Snell como una prueba de ajuste del modelo. Igualmente la Tabla 6 muestra el test de proporcionalidad global para el modelo final.

Tabla 7 Test de Proporcionalidad 2000 – Modelo Final

	rho	chi2	df	Prob > chi2
primaria	-0.0093	0.80	1	0.3720
secundaria	0.0051	0.24	1	0.6250
CFT	0.0010	0.01	1	0.9266
IP	-0.0105	1.01	1	0.3138
UNI	0.0159	2.33	1	0.1272
ln_yntl	-0.0003	0.00	1	0.9806
ln_edad	-0.0003	0.00	1	0.9812
s	0.0097	0.89	1	0.3457
Global		11.17	8	0.1922

Estimaciones Paramétricas

Se hicieron estimaciones paramétricas de máxima verosimilitud para la función Exponencial, Weibull y Log Logistic y Log Normal, siguiendo a Hosmer et Al. (1999). En esta sección sólo se consideraron los datos provenientes de CASEN 2006. Para poder construir las funciones de máxima verosimilitud - MV se asumió que todas las observaciones salieron del estado de desempleo al igual que en la sección anterior. Cabe recordar que esta es una falencia del seguimiento de análisis de supervivencia con datos de corte transversal. Igualmente, se usó el algoritmo de Newthton y Raphson para el proceso de maximización de la función de MV. Las funciones de supervivencia de cada una de las estimaciones paramétricas se presentan en la Tabla 8, además de la parametrización de los coeficientes de las co-variables y de los parámetros auxiliares

Tabla 8 Estimaciones Paramétricas y sus Funciones de Supervivencia

Distribution	Metric	Survivor function	Parameterization	Ancillary parameters
Exponential	AFT	$\exp(-\lambda_j t_j)$	$\lambda_j = \exp(-\mathbf{x}_j \beta)$	
Weibull	AFT	$\exp(-\lambda_j t_j^p)$	$\lambda_j = \exp(-p \mathbf{x}_j \beta)$	p
Lognormal	AFT	$1 - \Phi \left\{ \frac{\log(t_j) - \mu_j}{\sigma} \right\}$	$\mu_j = \mathbf{x}_j \beta$	σ
Loglogistic	AFT	$\{1 + (\lambda_j t_j)^{1/\gamma}\}^{-1}$	$\lambda_j = \exp(-\mathbf{x}_j \beta)$	γ

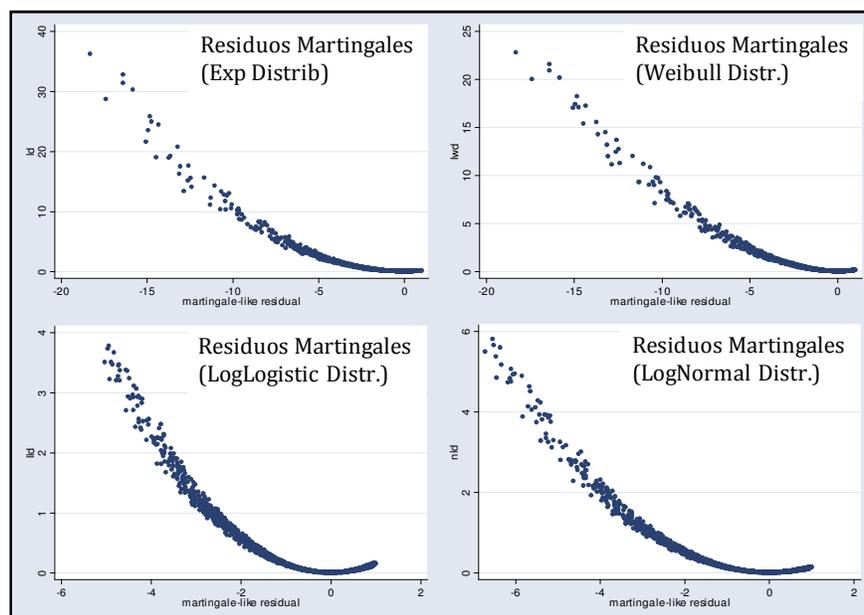
Fuente: Tomado de STATA Survival Analysis and Epidemiological Tables Reference Manual. Release 11. Pg. 358

Para la selección de un modelo paramétrico se mantuvieron las variables presentadas en la sección anterior para comparar y para evitar los efectos de interrelación que pudieran surgir de la imposición de una función específica, de hecho la significancia cambió para algunas variables en algunos modelos. Para la selección se analizaron los residuos martingales y Cox Snell, al igual que sus resultados de la maximización de la función de verosimilitud. A ese nivel se estimó la duración del desempleo para todas las funciones pero, por su aproximación se escogió el modelo log logístico.

Como el objetivo final es la estimación del tiempo promedio que un grupo de determinadas características estará en desempleo, se proveen las estimaciones del modelo seleccionado su *Accelerated Failure Time (AFT)*, y no las estimaciones de riesgos proporcionales, como el de Cox PHM, para poder ver el efecto directo de los coeficientes sobre la duración del desempleo.

Selección Modelo Paramétrico

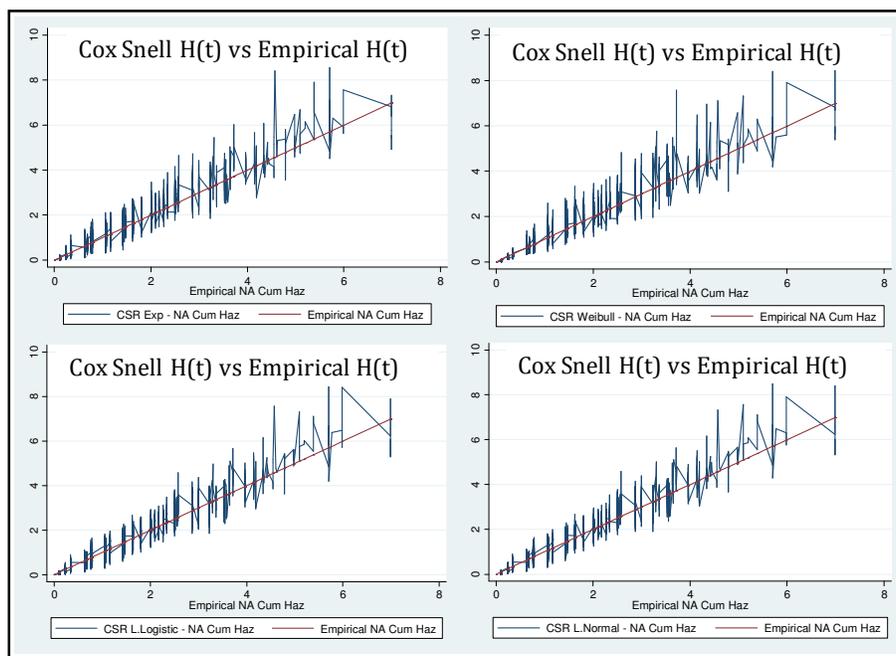
Figura 13 Desplazamiento de Verosimilitud y Residuos Martingales



La primera relación que se analizó fue la de los residuos martingales con la distancia de Cook o el estadístico de desplazamiento de verosimilitud. La comparación trata de ubicar el modelo cuyos rangos estén mayormente acotados para las dos variables alrededor de cero. La comparación sigue la metodología propuesta por Hosmer y Lemeshow (1999). Los modelos estimados para las distribuciones Exponencial y Weibull tienen los errores más dispersos al igual que el desplazamiento de verosimilitud. El modelo con distribución Log Logístico presenta la menor dispersión.

Otra apreciación de la bondad de ajuste de cada una de las estimaciones fue comparar la función riesgo estimada (Nelson – Alen) a partir de los errores Cox Snell con la función de riesgo empírica. La línea de 45° representa la función de riesgo empírica. Entre más alejada se encuentren los puntos correspondientes a las estimaciones paramétricas de la línea de 45° más dispersos son los errores de la estimación. Nuevamente son las distribuciones Log Logística y Log Normal las que presentan una distribución más cercana a la estimación empírica.

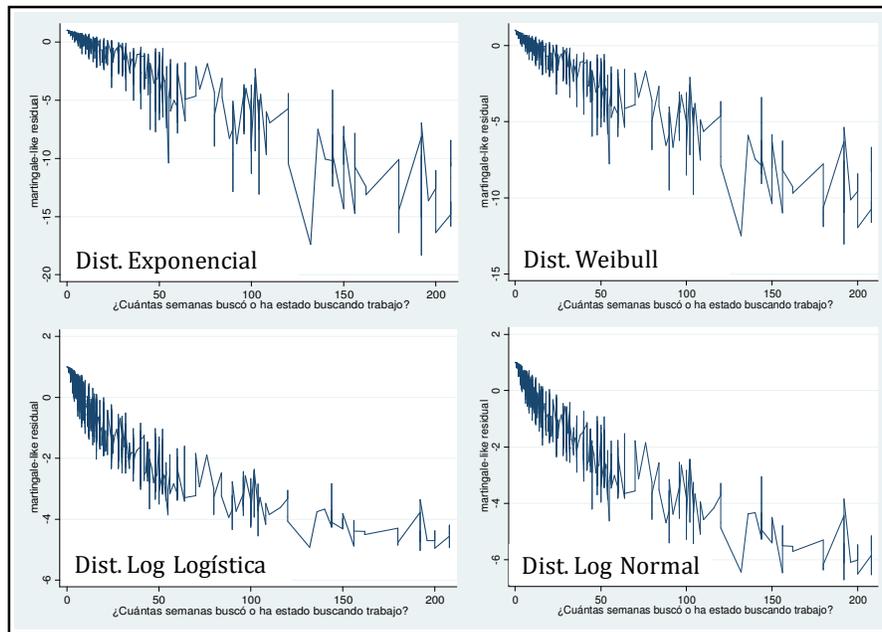
Figura 14 Estimado de la Función de Riesgo – Residuos Cox Snell (NA) vs Estimación Empírica (NA)



Finalmente se presenta la relación de los residuos con la duración del desempleo. A medida que la duración se aleja de la media los errores son muy dispersos. La distribución Log Logística tiene los mejores residuos para todos los rangos de la duración del desempleo. En ambos casos la estimación de los modelos correspondientes a una distribución Log Logística o a una distribución Log Normal, exhiben mejores resultados que en el caso de las otras dos estimaciones. Finalmente, se decidió por el modelo Log Logístico por tener su

desplazamiento de verosimilitud y sus residuos martingales más acotados que los del modelo que asume una distribución Log Normal y tener el máximo valor para la función de verosimilitud.

Figura 15 Residuos y Duración del Desempleo



Estimación de la Duración del Desempleo

Para la estimación de la duración del desempleo asumiendo una distribución Log Logística, tres de las variables no resultaron significativas al 10% de confianza y cuatro variables si el límite de la prueba se ajustaba a un 5%; a pesar de lo anterior, todos los efectos de las variables mantuvieron la misma dirección. Los resultados de las estimaciones se presentan en el Anexo A.3. La Tabla 9 presenta las estimaciones de la duración del desempleo para límites de la función de supervivencia de 17.5% a 22.5%.

Tabla 9 Estimación de la Duración del Desempleo (S=20%) - Reg. Log Logística

	D. Desempleo (S=22.5%)	D. Desempleo (S=20%)	D. Desempleo (S=17.5%)
Hombre Uni	24.04	26.36	29.17
Mujer Uni	24.62	27.00	29.87
Hombre IP	21.29	23.34	25.83
Mujer IP	21.80	23.91	26.45
Hombre CFT	18.55	20.33	22.50
Mujer CFT	18.99	20.83	23.04
Hombr Ed. Media	13.36	14.65	16.21
Mujer Ed. Media	13.68	15.00	16.60

Para la estimación de la Tabla 9 se asumieron las medias muestrales para el ingreso no laboral del individuo y del hogar, al igual que para el número de personas en el hogar. Se asumió que la persona característica vivía con sus padres y que el jefe de su hogar tenía 12 años de escolaridad. La función sistemática se evaluó sobre 27 años de edad. Las estimaciones indican que la duración del desempleo para la población con esas características se ubica entre 6 y 8 meses para los graduados de universidades (24 y 30 semanas).

Conclusiones

Se estimaron modelos de Cox PHM para dos bases de datos: CASEN 2006 y CASEN 2000, con el fin de analizar qué variables tienen qué grados de relación con la variable de duración del desempleo bajo las condiciones que cada base representa: la primera, un periodo de crecimiento moderado, y la primera un momento de salida de crisis. Efectivamente, se encuentra que durante la crisis la mayoría de variables socio demográficas (i.e. parentesco, número de personas, género, educación jefe de hogar) pierden significancia.

Igualmente, para el caso de la CASEN 2006 se encuentra que un mayor nivel educativo está asociado con un mayor tiempo buscando empleo. El anterior resultado se puede explicar con los costos que están dispuestos a asumir las personas con mayor nivel educativo para encontrar y negociar un buen trabajo. Por otro lado, usando datos de CASEN 2000, se encontró una relación similar en cada nivel de estudios con la excepción de los grados de primaria y de la población con educación universitaria. Lo anterior indicaría en cuanto a los graduados con educación básica o menos, que son los primeros en adaptarse al recibir menores salarios, probablemente por necesidad. En el caso de los profesionales, podría ser un indicio de que la población con niveles universitarios se adaptaría a cambios en el mercado laboral probablemente más rápido que personas con otros niveles educativos o que la demanda por profesionales no cayó tanto durante la crisis de 2000. Datos de otras recesiones en Chile podrían indicar que la población con grados universitarios salen más pronto del desempleo.

Debido a que algunas variables de los modelos usando Cox PHM resultaban significativas pero no se podía rechazar la hipótesis de proporcionalidad, para la estimación de las funciones de supervivencia en relación a las variables se hizo a través de estimaciones paramétricas, usando estimaciones de máxima verosimilitud para encontrar las estimaciones de las funciones de supervivencia Exponencial, Weibull, LogLogística y Log Normal. Se mantuvo la selección de variables de Cox PHM ya al ser un modelo no paramétrico se evaden efectos sobre las relaciones entre las variables que puedan surgir de asumir una forma funcional. Se encontró que los datos se ajustan mejor a la forma Log Logística y se desarrollaron las estimaciones puntuales para la duración del desempleo para todos los niveles de la educación superior y para la enseñanza media.

Referencias

- Cowan, Kevin; Micco, A.; Mizala, A.; Pagés, C.; Romaguera, P. (2003). "Un Diagnóstico del Desempleo en Chile". Documento encargado por el Ministerio de Economía al BID y a la U. de Chile. En: <http://www.dipres.cl>
- Fuenzalida, Marcelo y Ruiz-Tagle, Jaime (2008). "Households Financial Vulnerability and Financial Stability". Banco Central de Chile y Departamento de Economía U. de Chile. En: <http://www.microdatos.cl>
- Hosmer, David y Lemeshow, Stanley (1999). "Applied Survival Analysis: regression Modeling of Time to Event Data". Wiley Interscience Publications, New York.
- Martinez, Hermes (2003). "¿Cuánto duran los colombianos en el desempleo y en empleo?: Un Análisis de Supervivencia". Documento Cede No 35. Universidad de los Andes. Bogotá, Colombia.
- van Es, Bert; Klaasen, C.; Oudshoorn (2000). "Survival Analysis under cross-sectional sampling: length bias and multiplicative censoring". Journal of Statistical Planning and Inference No 91. Págs 295-312.
- Wooldridge, Jeffrey (2001). "Econometric Analysis of Cross Section and Panel Data". The MIT Press, Cambridge Massachusetts.

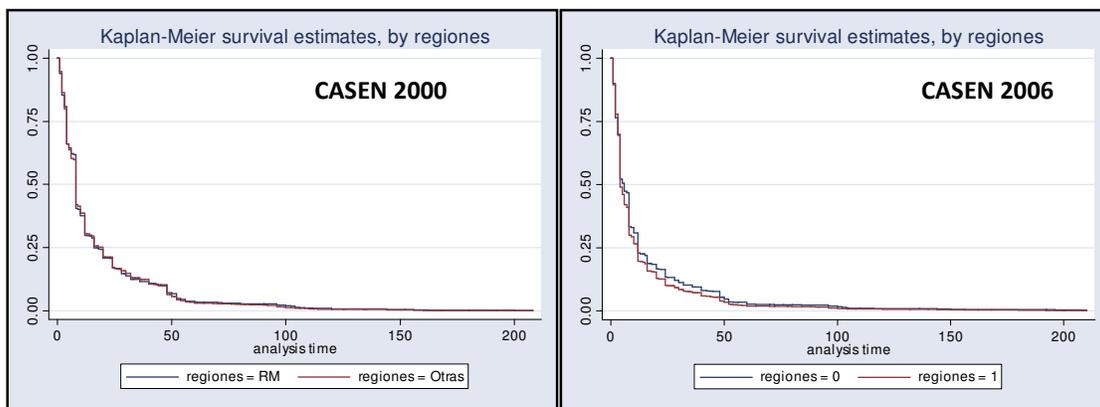
ANEXOS

A.1 Gráficas de Supervivencia para otras variables

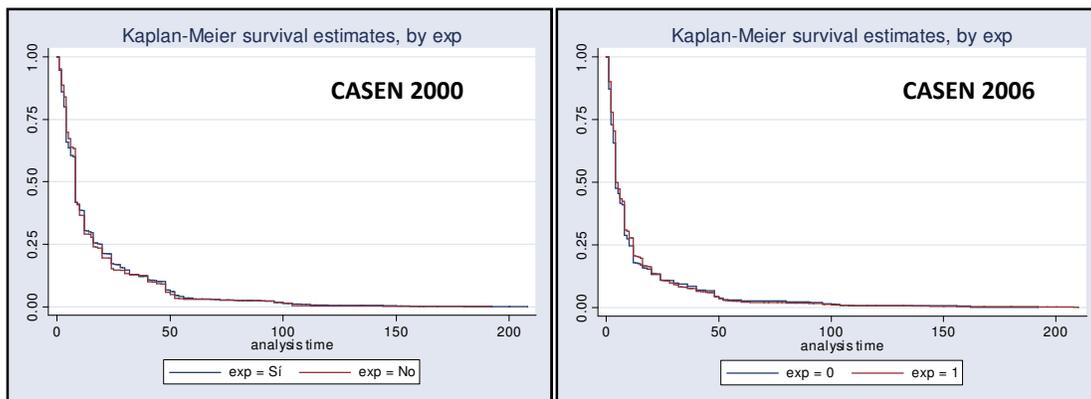
En esta sección se presenta el análisis gráfico de las funciones de supervivencia de otras variables consideradas en el análisis para la duración de búsqueda de empleo y la duración en el empleo. A las variables con demasiadas categorías como el parentesco con el jefe de hogar o el estado civil se omitió la presentación del análisis gráfico, pero se presentan igual las pruebas de ranking logarítmico en la sección de presentación de la base de datos.

¿Cuántas semanas buscó o ha estado buscando trabajo?

Regiones



Experiencia Previa



A.2 Construcción del Modelo Duración Desempleo

Estimaciones preliminares CASEN 2006

Las variables incluidas en el primer modelo provienen de la preselección hecha con los LR tests y con las regresiones individuales para las variables continuas presentadas en las Tablas 1 y 2.

Tabla A.1 Cox PHM - Todas las Variables

Number obs	6,843					
LR chi2(31)	306.69					
Prob > chi2	0.00					
No. of failures	6,843					
Time at risk	79,794.00					
Log likelihood	-54,111.95					
	Coef.	Std. Err	z	P> z	[95% Conf. Int]	
Espos(a) o Pareja	0.1196	0.0465	2.57	0.010	0.0284	0.2107
Hijo(a) ambos	-0.2035	0.0646	-3.15	0.002	-0.3302	-0.0769
Hijo(a) Jefe	-0.0553	0.0664	-0.83	0.405	-0.1854	0.0748
Hijo(a) Pareja	-0.1057	0.1025	-1.03	0.302	-0.3066	0.0951
Padre o Madre	-0.0790	0.2641	-0.30	0.765	-0.5966	0.4385
Suegro(a)	-0.2200	0.3829	-0.57	0.566	-0.9705	0.5305
Yerno o nuera	-0.1878	0.0900	-2.09	0.037	-0.3643	-0.0113
Nieto (a)	-0.1194	0.1325	-0.90	0.368	-0.3791	0.1403
Hermano(a)	-0.1001	0.1059	-0.95	0.345	-0.3077	0.1075
Cuñado(a)	0.0153	0.1572	0.10	0.922	-0.2927	0.3233
Otro Familiar	-0.1119	0.1095	-1.02	0.307	-0.3265	0.1028
No Familiar	0.0937	0.1344	0.70	0.486	-0.1698	0.3571
Union Libre	0.0402	0.0401	1.00	0.316	-0.0384	0.1187
Anulado	0.2760	0.3046	0.91	0.365	-0.3210	0.8730
Separado	0.0105	0.0646	0.16	0.870	-0.1161	0.1371
Divorciado	0.5287	0.4135	1.28	0.201	-0.2818	1.3392
Viudo	0.0896	0.1313	0.68	0.495	-0.1677	0.3469
Soltero	0.0200	0.0526	0.38	0.704	-0.0830	0.1230
sexo	-0.0702	0.0284	-2.48	0.013	-0.1258	-0.0146
regiones	0.0713	0.0298	2.39	0.017	0.0128	0.1297
primaria	-0.1015	0.0568	-1.79	0.074	-0.2128	0.0098
secundaria	-0.1378	0.0428	-3.22	0.001	-0.2218	-0.0539
CFT	-0.2043	0.1048	-1.95	0.051	-0.4097	0.0011
IP	-0.2326	0.0829	-2.80	0.005	-0.3952	-0.0701
UNI	-0.3228	0.0815	-3.96	0.000	-0.4825	-0.1631
In_ynl	0.0081	0.0030	2.72	0.007	0.0023	0.0139
In_nolbh	-0.0084	0.0041	-2.04	0.042	-0.0165	-0.0003
In_ytoth	0.0310	0.0112	2.77	0.006	0.0091	0.0528
educ_jefe	0.0071	0.0042	1.68	0.094	-0.0012	0.0153
In_edad	-0.5441	0.0537	-10.13	0.000	-0.6493	-0.4388
numper	0.0168	0.0075	2.26	0.024	0.0022	0.0315
yrs_educ2	-0.0034	0.0101	-0.33	0.741	-0.0232	0.0165

En negrilla variables no significativas a un nivel de confianza de 10%

De la estimación inicial se removieron todas las categorías de la variable Estado Civil y sólo se mantuvieron las dos categorías significativas de la variable Parentesco con el Jefe de Hogar. También se removió la variable de ingresos no laborales en el hogar y la de ingresos totales del hogar que perdían significancia al removerse las otras variables. Igualmente, se removió la variable años de educación primero que al incluirse con las variables dicotómicas para cada uno de los niveles de educación pierde significancia en

el modelo. Sin embargo, se usó la variable de años de educación sin las variables dicotómicas ya que las gráficas de los errores Schoenfeld de las variables dicotómicas para egresados de la educación secundaria, IPs o universidades (Figura A.1 y Figura A.2), resultan en un mayor problema de proporcionalidad que en el caso de los años de educación, pues los errores son más dispersos.

Figura A.1 Test de Proporcionalidad. Errores Schoenfeld Escalados – IPs y Universidades (Modelo reducido sin s)

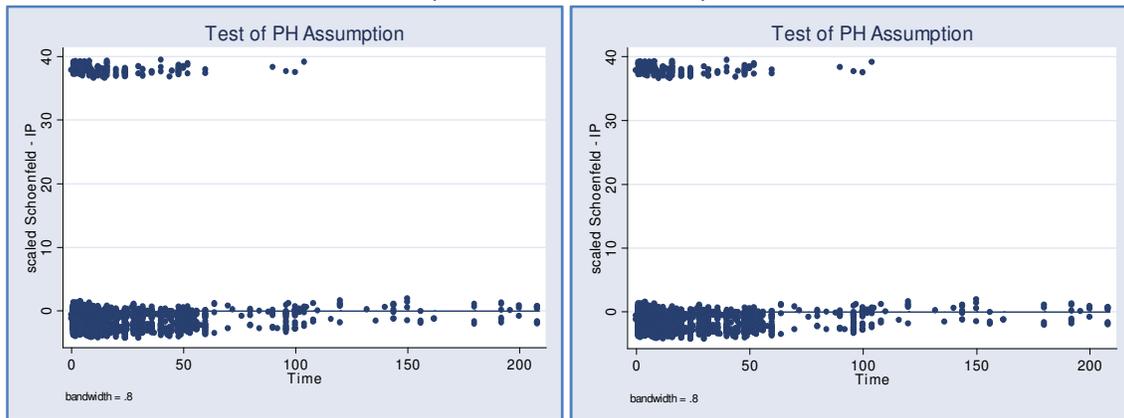
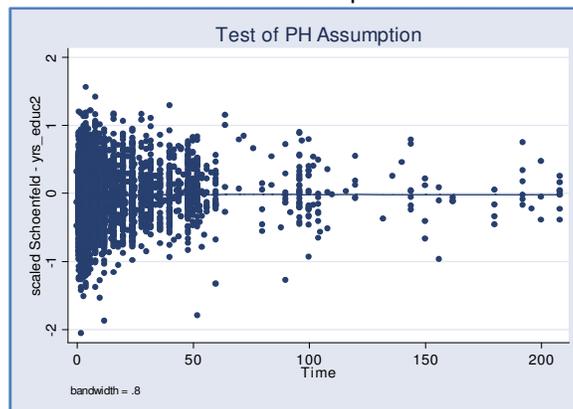


Figura A.2 Test de Proporcionalidad. Errores Schoenfeld Escalados – s (Modelo reducido sin variables dicotómicas para diferentes niveles educativos)



Posteriormente, en las pruebas de proporcionalidad se incluyeron en el modelo las interacciones de cada una de las variables con el logaritmo del tiempo ($\ln(t)$). Si alguna de las variables de interacción es significativa, la variable rompe el supuesto de proporcionalidad. La Tabla A.2 presenta el modelo final, que incluye variables que rompen el supuesto de proporcionalidad. Las variables no listadas no rompieron el supuesto de proporcionalidad.

Tabla A.2 Prueba de Proporcionalidad – Modelo Final

Number of obs	6,843					
LR chi2(17)	333.84					
Prob > chi2	0.00					
No. of failures	6,843					
Time at risk	79,794.00					
Log likelihood	-54,098.37					
	Coef.	Std. Err	z	P> z	[95% Conf. Int]	
Espos(a) o Pareja	0.1599	0.0372	4.30	0.000	0.087	0.233
Hijo(a)	-0.1256	0.0321	-3.91	0.000	-0.188	-0.063
primaria	-0.1218	0.0394	-3.09	0.002	-0.199	-0.045
secundaria	-0.1432	0.0305	-4.70	0.000	-0.203	-0.083
CFT	-0.2057	0.1022	-2.01	0.044	-0.406	-0.005
IP	-0.2318	0.0775	-2.99	0.003	-0.384	-0.080
UNI	-0.9128	0.1536	-5.94	0.000	-1.214	-0.612
sexo	0.1027	0.0466	2.20	0.028	0.011	0.194
regiones	0.0665	0.0295	2.25	0.024	0.009	0.124
ln_yml	0.0097	0.0028	3.48	0.001	0.004	0.015
ln_nolbh	-0.0057	0.0039	-1.43	0.152	-0.013	0.002
educ_jefe	0.0200	0.0062	3.24	0.001	0.008	0.032
ln_edad	-0.5074	0.0411	-12.33	0.000	-0.588	-0.427
numper	0.0167	0.0069	2.43	0.015	0.003	0.030
t						
UNI	0.2620	0.0590	4.44	0.000	0.146	0.378
sexo	-0.0966	0.0218	-4.44	0.000	-0.139	-0.054
educ_jefe	-0.0062	0.0029	-2.18	0.029	-0.012	-0.001

Estimaciones preliminares CASEN 2000

De los LR tests y las regresiones individuales para variables continuas se preseleccionaron las variables del modelo inicial presentado en la Tabla A.3. El parentesco con el jefe de hogar y el estado civil se sacaron del modelo final debido al escaso aporte explicativo de las variables categóricas independientes. Igualmente se procedió con la variable de educación del jefe del hogar y con la variable de género. Todas estas variables ya había tenido unas pruebas de Log Rank no significativas pero se consideraron en el modelo inicial por encontrar valores por debajo del 0.2 para el Log Rank.

En este caso solamente las variables de secundaria y universidad no rechazaron la hipótesis de no proporcionalidad. La variable de secundaria se mantuvo, porque es el principal efecto que busca el presente artículo, sin embargo su efecto no es significativo dado el tamaño de la varianza de su coeficiente que hace que los intervalos de confianza sean más amplios. Los resultados de la prueba de proporcionalidad final se presentan en la Tabla A.4, sólo los estimadores para los niveles educativos de Secundaria y Universitaria violan

el supuesto de proporcionalidad. Al ser parte de una variable categórica más amplia –niveles de educación– se mantuvieron en el modelo final.

Tabla A.3 Cox PHM – Todas las variables

Number obs	9,121
LR chi2(26)	298.61
Prob > chi2	0.00
No. of failures	9,121
Time at risk	144,815.00
Log likelihood	-74,710.42

	Coef.	Std. Err	z	P> z	[95% Conf. Int]	
Espos(a) o Pareja	0.0791	0.0422	1.87	0.061	-0.0036	0.1617
Hijo(a), Hijastro(a)	-0.1766	0.0440	-4.01	0.000	-0.2628	-0.0903
Padre o Madre	0.4650	0.3207	1.45	0.147	-0.1636	1.0935
Suegro (a)	0.2769	0.2708	1.02	0.307	-0.2539	0.8077
Yerno o nuera	-0.0338	0.0652	-0.52	0.604	-0.1615	0.0939
Nieto (a)	-0.1613	0.0756	-2.13	0.033	-0.3094	-0.0132
Hermano(a)	0.1329	0.0901	1.48	0.140	-0.0437	0.3095
Cuñado (a)	-0.0288	0.1185	-0.24	0.808	-0.2612	0.2035
Otro Familiar	-0.0126	0.0827	-0.15	0.879	-0.1747	0.1494
No Familiar	0.1307	0.1176	1.11	0.266	-0.0998	0.3612
Unión Libre	-0.0545	0.0365	-1.49	0.135	-0.1262	0.0171
Anulado	-0.0012	0.2518	0.00	0.996	-0.4947	0.4924
Separ. U. Legal	0.0479	0.0653	0.73	0.463	-0.0800	0.1758
Separ. U. de Hecho	0.0146	0.0725	0.20	0.840	-0.1274	0.1567
Viudo	-0.0394	0.0965	-0.41	0.683	-0.2285	0.1497
Soltero	-0.0203	0.0406	-0.50	0.617	-0.0999	0.0593
primaria	-0.0259	0.0437	-0.59	0.553	-0.1116	0.0598
secundaria	-0.0911	0.0366	-2.49	0.013	-0.1628	-0.0194
CFT	-0.1890	0.0844	-2.24	0.025	-0.3545	-0.0235
IP	-0.1364	0.0790	-1.73	0.084	-0.2912	0.0184
UNI	-0.0268	0.0843	-0.32	0.750	-0.1921	0.1384
sexo	-0.0183	0.0254	-0.72	0.471	-0.0680	0.0314
ln_ynl	0.0090	0.0055	1.62	0.106	-0.0019	0.0198
educ_jefe	-0.0020	0.0035	-0.58	0.561	-0.0089	0.0048
ln_edad	-0.5579	0.0440	-12.67	0.000	-0.6442	-0.4717
yrs_educ2	-0.0189	0.0082	-2.30	0.021	-0.0350	-0.0028

Tabla A.4 Prueba de Proporcionalidad – Modelo Final

Number of obs	9,121
LR chi2(10)	252.02
Prob > chi2	0.00
No. of failures	9,121
Time at risk	144,815.00
Log likelihood	-74,733.71

	Coef.	Std. Err	z	P> z	[95% Conf. Int]	
main						
primaria	-0.0214	0.0437	-0.4900	0.6250	-0.1070	0.0642
secundaria	-0.2054	0.0557	-3.6900	0.0000	-0.3144	-0.0963
CFT	-0.1861	0.0844	-2.2100	0.0270	-0.3515	-0.0208
IP	-0.1545	0.0789	-1.9600	0.0500	-0.3091	0.0001
UNI	-0.5330	0.1958	-2.7200	0.0060	-0.9168	-0.1492
ln_ynl	0.0099	0.0054	1.8200	0.0690	-0.0008	0.0205
ln_edad	-0.3763	0.0323	-11.6700	0.0000	-0.4395	-0.3131
yrs_educ2	-0.0193	0.0078	-2.4800	0.0130	-0.0345	-0.0041
tvc						
secundaria	0.05	0.02	2.33	0.02	0.01	0.08
UNI	0.20	0.07	2.87	0.00	0.06	0.34

A.3 Estimaciones Paramétricas

Regresión Log Logística

Tabla A.5 Coeficientes Estimación Log Logística

		Loglogistic regression -- accelerated failure-time form				
	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
No. of subjects	6843					
No. of failures	6843					
LR chi2(14)	416.77					
Prob > chi2	0					
Time at risk	79794.0003					
Log likelihood	-10312.773					
Esposo(a) o Pareja	-0.2160	0.0399	-5.41	0.0000	-0.2942	-0.1377
Hijo(a)	0.1362	0.0345	3.95	0.0000	0.0686	0.2037
primaria	0.1651	0.0418	3.95	0.0000	0.0832	0.2470
secundaria	0.1776	0.0322	5.52	0.0000	0.1146	0.2407
CFT	0.3279	0.1138	2.88	0.0040	0.1049	0.5510
IP	0.4659	0.0854	5.46	0.0000	0.2985	0.6332
UNI	0.5874	0.0763	7.70	0.0000	0.4379	0.7370
sexo	0.0239	0.0288	0.83	0.4070	-0.0325	0.0803
regiones	-0.0604	0.0322	-1.88	0.0610	-0.1235	0.0027
ln_ynl	-0.0134	0.0029	-4.60	0.0000	-0.0191	-0.0077
ln_nolbh	0.0044	0.0042	1.03	0.3030	-0.0039	0.0126
educ_jefe	-0.0145	0.0039	-3.72	0.0000	-0.0221	-0.0069
ln_edad	0.6012	0.0431	13.94	0.0000	0.5167	0.6857
numper	-0.0112	0.0073	-1.55	0.1210	-0.0255	0.0030
_cons	-0.4068	0.1713	-2.37	0.0180	-0.7425	-0.0711
/ln_gam	-0.4850	0.0101	-48.22	0.0000	-0.5047	-0.4652
gamma	0.6157	0.0062			0.6037	0.6280