

ANÁLISIS DE DURACIÓN EN EL CONSUMO DE CIGARRILLOS: RESULTADOS A PARTIR DE DATOS BRITÁNICOS

Andrew M. Jones

(University of York)

Nuria Badenes-Plá

(Universidad Complutense de Madrid)

INTRODUCCIÓN

En este trabajo presentamos evidencia acerca de los determinantes de la duración de las historias de consumo de cigarrillos a partir de los datos británicos que proporciona la encuesta de salud y estilos de vida (HALS: Health and Lifestyle Survey). Para ello se prueban distintas formas funcionales en modelos de duración paramétricos. Aparte de la consideración del gravamen que recae sobre el consumo de cigarrillos como una variable explicativa en la duración de las historias de los fumadores, se incluyen variables demográficas y una medida del consumo acumulado de cigarrillos en el pasado. Los resultados se presentarán para el total de la muestra utilizada, si bien se ha realizado además un análisis por cohortes de edad, que no se presenta aquí por razones de espacio.

El trabajo se organiza como sigue: tras la descripción de los datos de la encuesta en el primer epígrafe, exponemos brevemente los modelos econométricos utilizados tanto para la calcular la predicción del consumo de cigarrillos cuando esta información no está disponible, como los modelos de duración. Los resultados se exponen en el tercer epígrafe, para terminar con conclusiones en el cuarto. En el apéndice se da cuenta de las variables utilizadas y se exponen los resultados de las estimaciones.

1. DATOS

Los datos utilizados en este trabajo proceden del HALS, encuesta referida a una muestra representativa de individuos mayores de 18 años, que viven en hogares de Inglaterra, Escocia y Gales. La encuesta recoge la información obtenida a partir de dos visitas a los participantes. En la primera, se declara el estado de salud por parte de los entrevistados, y se responden preguntas relativas a dieta, deporte, consumo de alcohol y tabaco. También se pregunta acerca de la situación laboral, características demográficas, nivel de estudios y renta del hogar. En la segunda entrevista un/a enfermero/a toma

medidas sobre funciones cognitivas y físicas, y se proporciona un cuestionario que ofrece información sobre personalidad y salud mental.

Existen dos olas, elaboradas en 1984 y 1991 (este último año se entrevistó a los mismos individuos que siete años antes, para tratar de ofrecer una visión longitudinal, si bien de los 9003 participantes originales, en la segunda ola permanecieron 5352). Como nuestro interés se centra en conocer las historias pasadas de los fumadores actuales y ex-fumadores, y no existe un panel completo para todos los años en los que se ha fumado, construiremos un panel de forma artificial expandiendo los datos de la segunda ola según el número de años que se ha fumado hasta abandonar el consumo de cigarrillos (ex-fumadores), o hasta el momento de la encuesta (fumadores actuales). Ello significa que no contamos con el mismo número de observaciones para todos los individuos, pues la expansión de los datos dependerá del tiempo que lleven fumando o hayan estado fumando. Los datos se expanden con el fin de poder incorporar como variable explicativa la imposición sobre el tabaco, que varía en el tiempo. Al hacer esto, también seremos capaces de predecir para cada año el consumo esperado de cigarrillos, que luego acumularemos para calcular el "stock" de consumo pasado.

Los datos impositivos (introducidos en forma de logaritmo) fueron elaborados por Forster y Jones (2001). La recaudación por impuestos que gravan los cigarrillos se divide entre el valor de las ventas para construir el impuesto por cigarrillo, que se expresa en precios constantes de 1913-1914.

La necesidad de predicción del consumo de cigarrillos se produce porque en la encuesta utilizada, se conoce el consumo de cigarrillos de los actuales fumadores en un momento en el que todavía son fumadores, pero aquellos que han dejado de fumar ofrecen el dato relativo a la cantidad de cigarrillos que fumaba cuando decidieron dejar de fumar. Para la predicción utilizaremos el modelo de selección de Heckman, como se expone en el siguiente epígrafe.

Contamos con datos referidos a 5.166 individuos, y una vez expandidos, las observaciones se elevan a 128.905. El análisis de duración se realiza para la muestra total y por cohortes de edad¹. La distribución de fumadores y ex-fumadores por cohortes se presenta en la Tabla 1.

¹ Aunque por cuestiones de espacio no presentamos el detalle del análisis por cohortes, es relevante señalar que la división realizada es la siguiente: cohorte 1, para aquéllos nacidos antes de 1920, cohorte 2, para los nacidos entre 1920 y 1940, y cohorte 3, para los nacidos entre 1940 y 1970.

Tabla 1. Número de observaciones (muestra expandida) e individuos (muestra no expandida) correspondientes a fumadores y ex-fumadores por cohortes.

	Fumadores	Ex-fumadores	Total (no expandido)	Total (expandido)
Cohorte 1	318	577	895	36.684
Cohorte 2	960	816	1.776	56.347
Cohorte 3	1.660	835	2.495	35.874
<i>Total</i>	<i>2.938</i>	<i>2.228</i>	<i>5.166</i>	<i>128.905</i>

En la Tabla 2, se presenta, también por cohorte, la edad media de los individuos y la media de años de duración en el consumo de tabaco.

Tabla 2. Edad media y duración media del consumo de cigarrillos por cohortes.

	N	Edad media	Duración media
Cohorte 1	895	72,92	40,98
Cohorte 2	1.776	55,62	31,72
Cohorte 3	2.495	32,65	14,37

Las cantidades declaradas de consumo diario de cigarrillos entre los fumadores, se presentan en el siguiente histograma, que muestra que el consumo más frecuente (28,3%) es de un paquete diario (20 cigarrillos), ya que es en esta cuantía donde se sitúa la moda de la distribución.

■ ■ ■

Aunque nuestro interés no se centra en el estudio de los determinantes de la decisión de comenzar a fumar, comprobamos que como suele ocurrir, la edad de comienzo de consumo se concentra en torno a los 16 años. Otros descriptivos de interés se presentan en la Tabla 3.

Tabla 3. Número de individuos que deja de fumar por sexo y cohorte

		Mujeres	Hombres	Total
Cohorte=1	Fumadores	156	164	320
	Ex-fumadores	219	356	575
	<i>Total</i>	<i>375</i>	<i>520</i>	<i>895</i>
Cohorte=2	Fumadores	546	417	963
	Ex-fumadores	330	483	813
	<i>Total</i>	<i>876</i>	<i>900</i>	<i>1776</i>
Cohorte=3	Fumadores	894	775	1669
	Ex-fumadores	433	393	826
	<i>Total</i>	<i>1327</i>	<i>1168</i>	<i>2495</i>

Una variable crucial para nuestro análisis de duración es la cantidad de cigarrillos que se fuma al día. No la introduciremos directamente, pero la utilizaremos para calcular una

medida del stock de cigarrillos acumulado. Antes de predecir el consumo diario para los ex-fumadores, comprobamos que el perfil de consumo a lo largo del tiempo no es constante, ya que si fuese así, nos bastaría con la cantidad declarada y el número de años para calcular el stock de cigarrillos acumulado. Basándonos en la información de los fumadores, predecimos el consumo de cigarrillos a lo largo del tiempo, y comprobamos que para ninguna de las cohortes se observa un perfil plano. Después, utilizamos el modelo de selección de Heckman para predecir el consumo diario de cigarrillos, incluyendo por supuesto, el tiempo como una variable explicativa más. El stock de cigarrillos se introduce como una variable explicativa no a partir de las cantidades consumidas, sino como intervalos de consumo a través de variables dummy².

2. METODOLOGÍA

Modelo de selección de Heckman:

El modelo de selección de Heckman supone que existe la siguiente relación subyacente:

$$y_i = x_j\beta + u_{1j}$$

pero la variable dependiente (en nuestro caso, consumo diario de cigarrillos) no se observa siempre (en particular, para los ex-fumadores). La variable dependiente se observa para el individuo j si se cumple que:

$$z_j\gamma + u_{2j} > 0$$

donde

$$u_1 \approx N(0, \sigma)$$

$$u_2 \approx N(0, 1)$$

$$\text{corr}(u_1, u_2) = \rho$$

Modelos de duración:

Supongamos una variable aleatoria T , que describa la longitud temporal transcurrida hasta que ocurra cierto evento de interés (en nuestro caso, hasta que se deje de fumar). Podríamos describir la probabilidad de que T fuese menor que cierto valor t a través de su función de distribución ($F(t)$), y la derivada de ésta nos proporcionaría la función de densidad:

² Es importante destacar que el stock de cigarrillos acumulado no representa exactamente los cigarrillos fumados, sino esa cantidad dividida entre 365, ya que las cantidades predichas son diarias, que se multiplican por el número de años como fumador (no de días), pero a efectos de la interpretación, esto no es relevante, y más teniendo en cuenta que el stock se introduce a través de variables dicotómicas.

$$f(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{dt} = \frac{\partial F(t)}{\partial t}$$

La función de supervivencia ($S(t)$) proporciona la probabilidad de que una historia (tiempo en el que se fuma) dure t periodos o más:

$$S(t) = \Pr(T \geq t) = 1 - F(t)$$

La función de riesgo (o razón de fallo) ($h(t)$) especifica la tasa instantánea de fallo en un momento dado ($T=t$) condicionado a que se haya sobrevivido hasta ese momento t . (Es decir, la probabilidad de dejar de fumar en un momento determinado, teniendo en cuenta que hasta entonces se ha sido fumador)

$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t)}{dt} = \frac{f(t)}{S(t)}$$

La función de razón de fallo también se puede expresar como

$$h(t) = -\frac{d \log S(t)}{dt}$$

e integrando, tenemos que

$$S(t) = \exp\left(-\int_0^t h(s) ds\right) = \exp(-\Lambda(t))$$

Donde $\Lambda(t)$ se conoce como función de riesgo o razón de fallo integrada. La forma de la función de riesgo caracteriza el proceso, de manera que si la pendiente de ésta es positiva (negativa) el proceso exhibe una dependencia positiva (negativa) de la duración.

El análisis de duración se puede abordar sin suponer nada acerca de la forma funcional de los datos que generan el proceso (enfoque no paramétrico) o asumiendo ciertos supuestos (enfoques semiparamétrico y paramétrico). Aunque presentaremos las

estimaciones Kaplan-Meier de la función de supervivencia, nuestro interés se centra en los resultados de modelos paramétricos, suponiendo distintas formas funcionales y eligiendo el mejor ajuste.

Existen dos posibilidades de ajuste de las funciones de supervivencia incluyendo el efecto de las variables independientes, modelo acelerado de fallo (accelerated failure time model, AFT) y modelos de tasa de riesgo proporcional (proportional hazard rate, PH). En los modelos AFT, el logaritmo del tiempo de supervivencia ($\ln t$) se puede expresar como una función lineal de los regresores, es decir:

$$\ln t_j = x_j \beta + u_j$$

donde x_j es el vector de regresores, β es un vector de coeficientes, y u_j es el término de error, con una función de densidad $f(\cdot)$ que determina el modelo de regresión. El efecto en los modelos AFT es un cambio en la escala temporal mediante el factor $\exp(-x_j \beta)$. En los modelos PH, los regresores presentan un efecto multiplicativo sobre la función de riesgo:

$$h(t_j) = h_0(t)g(x_j)$$

donde $h_0(t)$ se conoce como razón de fallo básico que recoge la heterogeneidad individual no contenida en las variables explicativas del modelo, y $g(x_j)$ es una función no negativa de los regresores. Los modelos exponencial y Weibull se pueden implementar como PH o AFT, mientras que el modelo Gompertz solamente se puede especificar como PH, y el modelo Gamma como AFT. Dependiendo de la función de supervivencia asumida, trabajaremos con distintos modelos. A continuación, resumimos en la Tabla 4 posibles formas de la función de supervivencia.

Tabla 4. Funciones de supervivencia para distintas distribuciones

Distribución	Medida	Función de Supervivencia $S(t)$
Exponencial	PH	$\exp[-(e^{x_j \beta})t_j]$
	AFT	$\exp[-(e^{-x_j \beta})t_j]$
Weibull	PH	$\exp[-(e^{x_j \beta})t_j^p]$
	AFT	$\exp[-(e^{-x_j \beta p})t_j^p]$

Gompertz	PH	$\exp\left[-\frac{e^{x_j\beta}}{\gamma}(e^{\gamma t_j} - 1)\right]$
Gamma (k>0)	AFT	$1 - I(\gamma, v)$ siendo $\gamma = k ^{-2}$ $v = \gamma \exp(k u)$
Gamma (k=0)	AFT	$1 - \phi(u)$ siendo $\phi(u)$ la función de distribución acumulada normal
Gamma (k<0)	AFT	$I(\gamma, v)$

Nota: cuando el parámetro de dependencia temporal p (γ) es mayor que 1 (positivo), el riesgo de fallo crece a medida que pasa el tiempo, y cuando es menor que 1 (negativo), decrece. El modelo exponencial está anidado dentro del modelo Weibull (cuando $p=1$) y dentro del modelo Gompertz (cuando $\gamma=0$), y el modelo Weibull está anidado en la especificación Gamma, cuando $k=1$

3. RESULTADOS

Comenzaremos exponiendo los resultados de las predicciones obtenidas según el modelo de selección de Heckman. Los resultados de las regresiones por cohortes, se muestran en el apéndice. Simplemente destacaremos el efecto de algunas variables. En cuanto a la selección de los individuos como fumadores, los resultados muestran una probabilidad de selección siempre mayor cuando se convive con otros fumadores, para cualquiera de las cohortes. Ser hombre, y casado se asocia a una menor probabilidad de selección como fumador, así como padecer alguna discapacidad. En cuanto a la cantidad de cigarrillos fumada diariamente, es destacable que convivir con otros fumadores se asocia siempre con consumos más elevados de cigarrillos. En las cohortes mayores, la edad en la que se comienza a fumar presenta un significativo efecto negativo: es decir, cuanto más joven se comenzó a fumar, mayor es el consumo diario. La situación conyugal en general sigue un patrón común, de manera que divorciados y separados fuman más que aquellos que continúan casados.

Puesto que en nuestras estimaciones ρ no es cero³ la estimación por MCO daría lugar a resultados sesgados. El histograma de las predicciones Heckman para el consumo diario de cigarrillos muestra el siguiente perfil:

▲

Una vez que se ha predicho el consumo diario de cigarrillos, podemos calcular una medida acumulada del stock consumido año tras año. La predicción es diferente

³ Obtuvimos valores de rho 0.45060, 0.39505 y -0.15863, para las cohortes primera, segunda y tercera respectivamente.

para cada año, y se agrega por los años que los individuos declaran ser fumadores⁴. Esta variable se introduce en forma dicotómica según se pertenezca o no a uno de los cinco intervalos en los que se ha categorizado. En la Tabla 5 se resume la información del porcentaje de observaciones que pertenece a cada tramo para el total de la muestra y para el último año como fumadores.

Tabla 5. Distribución porcentual de los datos expandidos o no por pertenencia a intervalo de stock acumulado de cigarrillos

Fagday € intervalo	Total observaciones	Ultimo año como fumadores
Interv1: [0,100)	31,90%	19,11%
Interv2: [100, 200)	16,15%	15,87%
Interv3: [200, 300)	12,10%	12,54%
Interv4: [300, 400)	9,70%	10,03%
Interv5: [400 o más)	30,15%	42,45%

Además de introducir el stock de cigarrillos consumidos en la especificación paramétrica de modelos de duración, nos interesa comprobar la relevancia de otras variables, como el efecto de los impuestos sobre cigarrillos, la convivencia con otros fumadores, y una serie de variables demográficas que se explican en el apéndice.

Los resultados para la muestra total y por cohortes de distintas especificaciones funcionales, da lugar a los valores (en logaritmos) de la función de máxima verosimilitud que se presentan en la Tabla 6.

Tabla 6. Valor del logaritmo de la función de máxima verosimilitud.

	Whole	Cohort=1	Cohort=2	Cohort=3
Exponential	-4338,9882	-842,57341	-1487,1792	-1923,3935
Weibull	-4298,1777	-835,28955	-1483,473	-1895,8669
Weibull+fr(gamma)	-4297,7878	-835,28957	-1483,473	-1895,867
Gompertz	-4316,4501	-828,95771	-1474,4523	-1882,525
Gompertz+fr(gamma)	-4305,9217	-828,92216	-1473,7706	-1879,0929
Gamma	-4298,0623	-829,8456	-1481,7407	-1894,9277

Por coherencia en la interpretación de resultados para la muestra total y el análisis desagregado por cohortes, se escogió la presentación de los resultados asumiendo una distribución Gompertz con heterogeneidad no observable (que se modeliza según una distribución Gamma). La existencia de heterogeneidad no

⁴ Como la cantidad predicha es de consumo diario de cigarrillos, al acumular por años, no contamos con el stock real, sino dividido entre 365, por lo que no deben extrañar las pequeñas cuantías acumuladas.

observable significa que existe información de la que no disponemos que afecta de forma diferente sobre los individuos. La heterogeneidad se incorpora de forma multiplicativa sobre la función de razón de fallo:

$$h(t | \alpha) = \alpha h(t)$$

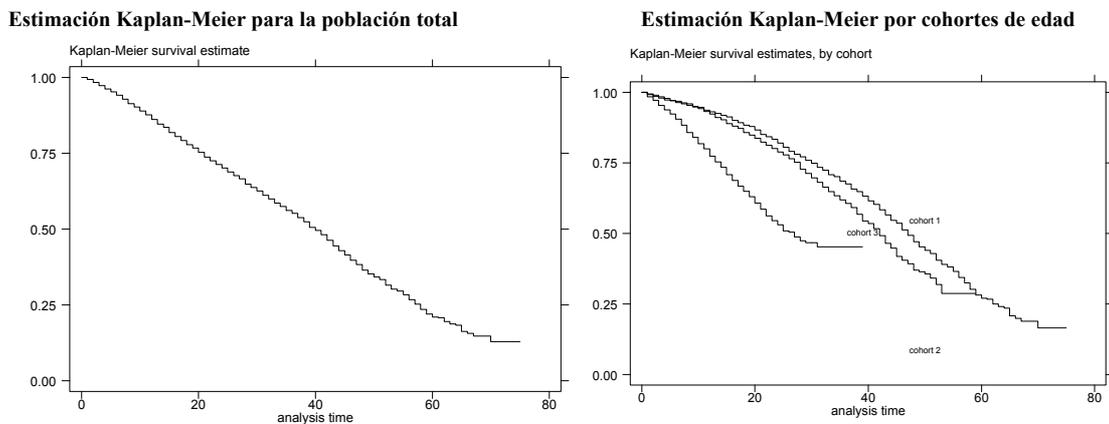
donde α es el factor que incorpora la heterogeneidad y suponemos que es una cantidad aleatoria y positiva con media unitaria y varianza igual a θ (por razones de identificabilidad). Si la varianza estimada resultase nula, contaríamos con un modelo sin heterogeneidad no observable.

Los resultados de la regresión se presentan en el apéndice, y es necesario recalcar, que un coeficiente igual a la unidad significa efecto nulo del regresor sobre el riesgo de abandono (en este caso, de dejar de fumar). Si el coeficiente es mayor que la unidad, ello implica que esa variable influye positivamente sobre la probabilidad de dejar de fumar (negativamente sobre la duración del hábito), y si el coeficiente es menor que la unidad, la variable incide negativamente sobre el "riesgo" de abandono de la condición de fumador.

La variable impositiva presenta el efecto esperado, es decir, que impuestos más elevados hacen disminuir la duración de las historias como fumadores, y de forma muy significativa. La magnitud del coeficiente de *Intax* es 2,15 y ello significa que un incremento de una unidad en la variable multiplicaría por 2,15 el riesgo de fallo (es decir, dejar de fumar). La variable más significativa de todas es la convivencia con otros fumadores, que incide negativamente sobre el riesgo de abandono, o alarga los periodos como fumador. La magnitud del efecto es multiplicar por 0,4 el riesgo de fallo. Ser hombre implica mayor duración en el consumo de tabaco frente a ser mujer, como muestra un coeficiente ligado a *male* de 0,11. En cuanto al stock acumulado de cigarrillos, se observa un efecto gradiente. Aquellos individuos que acumulan un número de cigarrillos mayor, presentan menor probabilidad de abandonar el hábito, lo que confirma el carácter adictivo del consumo de cigarrillos. Con respecto a la variable excluida, *interv5*, cuyo coeficiente asociado sería unitario, la pertenencia a intervalos de menor stock acumulado incrementa sucesivamente el riesgo de fallo hasta prácticamente duplicarlo para aquellos que se sitúan en el intervalo de menor stock (el coeficiente asociado a *interv1* es 1,98). En cuanto a niveles de estudios, pertenecer al colectivo sin estudios implica duraciones más prolongadas, y contar con un nivel universitario, lo contrario.

El valor estimado de la varianza, $\theta=0,434$ indica que efectivamente existe heterogeneidad no observable, y al ser positivo γ (igual a 0,041), el modelo exhibe dependencia temporal positiva. Si γ fuese nulo, estaríamos ante el modelo exponencial, pero como muestran los valores del logaritmo de la función de máxima verosimilitud, el modelo de riesgo independiente del tiempo no presenta el mejor ajuste).

Las estimaciones no paramétricas de la función de supervivencia a partir del estimador de Kaplan-Meier, nos proporcionan una guía que pone de manifiesto las diferencias entre el comportamiento teniendo en cuenta la población total, y separando por cohortes.



En las estimaciones no paramétricas de la función de supervivencia se aprecia que el "fallo" no se produce en todo caso, es decir, hay individuos que nunca dejan de fumar, ya que la función no desciende hasta cero. La estimación de la probabilidad de sobrevivir como fumador muestra que ésta decrece con el tiempo. Al considerar cohortes de más edad, la función toma valores mínimos más pequeños, ya que cuanto más largas son las historias de los fumadores, es menos probable encontrar fumadores que no abandonan nunca el hábito. Y como es obvio, las cohortes mayores presentan duraciones como fumadores más largas simplemente porque el tiempo en consideración es mayor.

Los gráficos de la estimación paramétrica (según una distribución Gompertz con heterogeneidad no observable) de la función de supervivencia y razón de fallo son las siguientes:

Función de supervivencia

Función de razón de fallo



Tal como se ha expuesto en el comentario de resultados, la probabilidad de sobrevivir en el estado "fumador" presenta una dependencia negativa del tiempo, o dicho de otra

forma, el riesgo de abandonar la condición de fumador, crece a medida que se fuma durante más años. Los gráficos están calculados para el valor medio de los parámetros, pero si se desea conocer la influencia de cada uno de ellos, se puede consultar el apéndice.

4. CONCLUSIONES

Los efectos del consumo de tabaco han sido ampliamente constatados como nocivos (véase Chaloukpa y Warner (2000)), pero muchos fumadores que desearían dejar de fumar no consiguen hacerlo por su carácter adictivo. El consumo adictivo de tabaco, presenta una oportunidad de recaudación para el sector público por la inelasticidad de las demandas de los fumadores, por lo que de alguna manera, al sector público le conviene que haya fumadores. En este trabajo se ha puesto de manifiesto que si el objetivo del gobierno es proteger al fumador -teniendo en cuenta que la salud es un bien preferente- un posible instrumento es el gravamen de cigarrillos. Es decir, por un lado gravar el consumo de cigarrillos llena las arcas públicas, pero al mismo tiempo se reduce el número de contribuyentes, ya que los impuestos sobre el tabaco inciden positivamente sobre la decisión de dejar de fumar. El análisis de duración que aquí hemos presentado no se centra exclusivamente en el efecto del gravamen de cigarrillos. Una novedad importante en este tipo de trabajos es la inclusión de una medida del stock de cigarrillos acumulado a lo largo del tiempo como variable explicativa en la duración de las historias de los fumadores. Tal y como se intuía a priori, cuanto mayor es el stock acumulado de cigarrillos, más difícil es dejar de fumar, lo que prueba el carácter adictivo de este consumo. A lo largo de todo el trabajo, una variable sorprendentemente significativa en las decisión de dejar de fumar, es la convivencia en el hogar con otros fumadores: es mucho más difícil dejar de fumar si más personas fuman en casa. El nivel de educación también aparece como variable significativa en el análisis de duración. Los más educados fuman por menos tiempo que los menos educados, a pesar de la correlación existente entre nivel de estudios y la renta, (y teniendo en cuenta además el elevado precio del tabaco en el Reino Unido), lo que revela un mayor nivel de concienciación entre los más educados. Para el análisis de datos agregados, los hombres aparecen con duraciones más prolongadas que las mujeres.

Cuando se analizan separadamente cohortes de edad, destaca el hecho de que la influencia de la convivencia con más fumadores es más relevante para las generaciones más jóvenes. Además al extenderse el consumo de cigarrillos entre las mujeres a medida

que pasa el tiempo, ser hombre deja de ser relevante en la explicación de la duración del consumo de tabaco.

Analizados los determinantes de la duración en el consumo de cigarrillos, queda patente que a parte de la exigencia de gravámenes cada vez mayores sobre los cigarrillos, los fumadores presentan capacidad de respuesta ante otras variables como la concienciación, por ejemplo. Lo que no está tan claro es que a los gobiernos les convenga utilizar otras medidas que no supongan aumento en sus ingresos. Es necesario tener en cuenta que un fumador es, o acabará siendo a la larga, un adicto en la mayoría de los casos. En el pecado va la penitencia, y los fumadores asumen un riesgo sobre su salud, que bien podría ser minorado de forma que no implicase además pagar más y más impuestos.

APÉNDICE

Descripción de las variables.

área de residencia: (=1 si reside en cada una de las siguientes áreas) **wales north nwest yorks wmids emids anglia swest seast london scot**

acupast: cantidad predicha del consumo pasado acumulado de cigarrillos

cohort::
=1 nacidos antes de 1920
=2 nacidos entre 1920 y 1940
=3 nacidos entre 1940 y 1970

dis: (=1) si presenta discapacidad

ownh: estado de salud declarado (=1: excelente, =2: bueno, =3: normal, =4: malo)

male: (=1 si es hombre)

(máximo nivel de estudios alcanzado)

lhqnone: ninguno

lhqO: nivel CSE/O

lhqA: nivel A

lhqhnd: por debajo de licenciatura

lhqdg: licenciatura

lhqoth: otro o no conocido

(origen étnico)

ethbawi: negro, africano

etheur: blanco/europeo

ethipb: indio, pakistaní, bangladeshi

ethothnw: otro origen no blanco

ethmiss: desconocido

parsm1: (=1) solo la madre fumadora

parsm2: (=1) solo el padre fumador

parsm3: (=1) ambos fumadores

quit: (=1) si dejó de fumar

agestrt: edad a la que comenzó a fumar

smkyrs2: número de años como fumador

T: variable temporal según año en el que se fuma. (TN es la potencia N-ésima)

smother: (=1) convivencia con otros fumadores en el hogar

survage: edad en el momento de la encuesta

(estatus laboral)

full: (=1) tiempo completo

part: (=1) tiempo parcial

retd: (=1) jubilado

stdnt: (=1) estudiante

unemp: (=1) parado

(estado marital)

widow: (=1) viudo

divorce: (=1) divorciado legalmente

single: (=1) soltero

seprd: (=1) casado pero separado de la pareja

married: (=1) casado viviendo con la pareja

Modelo de selección de Heckman. Cohorte=1

```

Heckman selection model -- two-step estimates      Number of obs      =      895
(regression model with sample selection)          Censored obs       =      577
                                                    Uncensored obs     =      318

                                                    Wald chi2(14)      =      249.87
                                                    Prob > chi2        =      0.0000
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

fagday						
T	-.0463203	.0958856	-0.48	0.629	-.2342527	.141612
agestrt	-.2050559	.0879027	-2.33	0.020	-.3773419	-.0327698
ownh	-.4979916	.4138658	-1.20	0.229	-1.309154	.3131704
divorce	8.784768	3.762787	2.33	0.020	1.409842	16.15969
single	1.987509	1.740368	1.14	0.253	-1.423549	5.398568
part	3.700527	2.049153	1.81	0.071	-.3157385	7.716793
parsm3	2.840523	1.349142	2.11	0.035	.1962533	5.484792
wmids	3.96894	1.604336	2.47	0.013	.8244991	7.11338
anglia	-7.791788	3.092464	-2.52	0.012	-13.85291	-1.730669
scot	2.36067	1.332982	1.77	0.077	-.2519264	4.973266
lhqnone	1.639606	1.345875	1.22	0.223	-.9982595	4.277472
lhqdg	4.76394	2.304623	2.07	0.039	.2469625	9.280917
lhqoth	3.507438	2.056792	1.71	0.088	-.5237999	7.538676
_cons	15.81405	6.935898	2.28	0.023	2.219944	29.40816

select						
T	.0640472	.004506	14.21	0.000	.0552156	.0728788
dis	-.3415028	.1019401	-3.35	0.001	-.5413016	-.1417039
smother	.455613	.0873011	5.22	0.000	.2845061	.62672
male	-.7438158	.1085767	-6.85	0.000	-.9566223	-.5310094
_cons	-2.751769	.2115691	-13.01	0.000	-3.166436	-2.337101

mills						
lambda	3.553261	1.703747	2.09	0.037	.2139788	6.892543

rho	0.45060					
sigma	7.8856567					
lambda	3.553261	1.703747				

Modelo de selección de Heckman. Cohorte=2

```

Heckman selection model -- two-step estimates      Number of obs      =      1776
(regression model with sample selection)          Censored obs       =      816
                                                    Uncensored obs     =      960

                                                    Wald chi2(23)      =      566.44
                                                    Prob > chi2        =      0.0000
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

fagday						
T	3.284162	1.888894	1.74	0.082	-.4180014 6.986326	
T2	-16.5317	9.643341	-1.71	0.086	-35.4323 2.368901	
T3	345.5841	205.4637	1.68	0.093	-57.11743 748.2856	
T4	-25373.37	15576.53	-1.63	0.103	-55902.8 5156.059	
smother	2.128065	.6007423	3.54	0.000	.9506319 3.305498	
agestrt	-.1509073	.0810041	-1.86	0.062	-.3096724 .0078577	
ownh	.7891221	.3111684	2.54	0.011	.1792433 1.399001	
seprd	5.249275	1.728138	3.04	0.002	1.862186 8.636364	
full	2.00054	.6506639	3.07	0.002	.7252627 3.275818	
retdd	-2.238857	.9284415	-2.41	0.016	-4.058569 -.4191449	
nwest	2.077163	.8080818	2.57	0.010	.4933518 3.660974	
anglia	-3.663075	1.637064	-2.24	0.025	-6.871662 -.4544883	
scot	3.202434	.8612408	3.72	0.000	1.514433 4.890435	
lhqnone	-1.321985	.7491766	-1.76	0.078	-2.790344 .1463744	
lhqA	-2.828963	1.609783	-1.76	0.079	-5.984079 .3261531	
lhqdg	-3.965819	1.359532	-2.92	0.004	-6.630452 -1.301185	
_cons	-8.380722	13.94366	-0.60	0.548	-35.7098 18.94835	

select						
T	.0790466	.0040779	19.38	0.000	.071054 .0870392	
smother	.7561309	.0739211	10.23	0.000	.6112481 .9010137	
male	-.7653944	.0871689	-8.78	0.000	-.9362424 -.5945465	
ethipb	1.838884	.4219583	4.36	0.000	1.011861 2.665907	
ethbawi	1.279939	.4778155	2.68	0.007	.3434378 2.21644	
lhqnone	-.178749	.0851643	-2.10	0.036	-.3456679 -.0118301	
lhqdg	-.3366636	.1419478	-2.37	0.018	-.6148762 -.058451	
nwest	-.15809	.1058601	-1.49	0.135	-.365572 .0493921	
anglia	-.5595727	.1895831	-2.95	0.003	-.9311487 -.1879967	
swest	-.3614283	.1350478	-2.68	0.007	-.626117 -.0967395	
seast	-.1850924	.1052464	-1.76	0.079	-.3913715 .0211867	
london	-.2372726	.1189684	-1.99	0.046	-.4704464 -.0040987	
married	-.4851648	.0950992	-5.10	0.000	-.6715559 -.2987737	
full	.4362538	.0901253	4.84	0.000	.2596116 .6128961	
part	.3294469	.122246	2.69	0.007	.0898492 .5690446	
unemp	.5546066	.1690805	3.28	0.001	.2232149 .8859982	
_cons	-2.007757	.1738384	-11.55	0.000	-2.348474 -1.66704	

mills						
lambda	3.521507	1.342757	2.62	0.009	.8897524 6.153262	

rho	0.39505					
sigma	8.9140445					
lambda	3.521507	1.342757				

Modelo de selección de Heckman. Cohorte=3

```

Heckman selection model -- two-step estimates      Number of obs      =      2495
(regression model with sample selection)          Censored obs       =      835
                                                    Uncensored obs     =      1660

                                                    Wald chi2(27)      =      739.83
                                                    Prob > chi2        =      0.0000
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

fagday						
T	1.281943	.2835253	4.52	0.000	.7262436 1.837643	
T2	-5.732916	1.77418	-3.23	0.001	-9.210246 -2.255586	
T3	96.97322	33.84031	2.87	0.004	30.64743 163.299	
smother	.6015064	.5038631	1.19	0.233	-.3860472 1.58906	
ownh	1.376223	.2498617	5.51	0.000	.8865036 1.865943	
divorce	-1.82136	1.191551	-1.53	0.126	-4.156757 .514036	
married	-2.401064	1.058705	-2.27	0.023	-4.476088 -.3260392	
single	-2.562648	1.068283	-2.40	0.016	-4.656444 -.4688518	
full	2.172028	.471839	4.60	0.000	1.24724 3.096815	
unemp	1.862677	.6958778	2.68	0.007	.4987818 3.226573	
wmids	-1.39735	.6611236	-2.11	0.035	-2.693128 -.1015714	
seast	-1.850777	.5665731	-3.27	0.001	-2.96124 -.7403144	
london	-1.362833	.6470375	-2.11	0.035	-2.631003 -.0946626	
lhqnone	-15.97296	7.9705	-2.00	0.045	-31.59485 -.3510645	
lhqO	-16.63458	7.97794	-2.09	0.037	-32.27105 -.9981025	
lhqA	-15.69427	8.021672	-1.96	0.050	-31.41646 .027915	
lhqhnd	-17.23777	7.988266	-2.16	0.031	-32.89449 -1.581058	
lhqdg	-17.61182	7.993225	-2.20	0.028	-33.27825 -1.945389	
lhqoth	-16.69945	8.014718	-2.08	0.037	-32.40801 -.9908932	
parsm3	1.89802	.3967289	4.78	0.000	1.120445 2.675594	
_cons	21.29416	8.317899	2.56	0.010	4.991373 37.59694	

select						
T	.0779753	.0045999	16.95	0.000	.0689596 .0869909	
dis	-.2358355	.0696163	-3.39	0.001	-.372281 -.0993901	
smother	.5520009	.0528022	10.45	0.000	.4485104 .6554914	
ethipb	.9184968	.3518777	2.61	0.009	.2288292 1.608164	
lhqO	.1801001	.0686786	2.62	0.009	.0454926 .3147076	
lhqdg	-.3246283	.0912299	-3.56	0.000	-.5034355 -.145821	
scot	.3561493	.0996494	3.57	0.000	.16084 .5514586	
married	-.6501037	.1054078	-6.17	0.000	-.8566992 -.4435082	
single	.4044839	.1240693	3.26	0.001	.1613124 .6476553	
full	-.2315303	.0665306	-3.48	0.001	-.3619279 -.1011327	
part	-.1840924	.0948689	-1.94	0.052	-.3700321 .0018473	
_cons	-.377107	.1259563	-2.99	0.003	-.6239768 -.1302373	

mills						
lambda	-1.266906	1.79812	-0.70	0.481	-4.791156 2.257345	

rho	-0.15863					
sigma	7.9865539					
lambda	-1.2669056	1.79812				

BIBLIOGRAFÍA

Chaloukpa, F. J., and K. Warner, (2000), "The Economics of smoking" (1540-1627), in *Handbook of Health Economics*, Volume 1. Edited by A. Cuyler and J.P. Newhouse. Elsevier Science B.V.

Douglas, S. M. (1998), "The duration of smoking habit", *Economic Inquiry*, 36(1):49-64.

Forster, M. and A. M. Jones (2001), "The role of tobacco taxes in starting and quitting smoking: duration analysis of British data" *Journal of the Royal Statistics Society* 164(3):517-547.

Jones, A.M. (1994), "Health, addiction, social interaction and the decision to quit smoking" *Journal of Health Economics* 13:93-110.

Jones, A. M. and J. Posnett (1988), "Revenue and welfare effects of cigarette taxes" *Applied Economics* 1988 (20):1223-32.

López, A. (2002) "How important are tobacco prices in the propensity to start and quit smoking". An analysis of smoking histories from the Spanish National Health Survey. *Health Economics* (forthcoming).

Neumann, G.R. (1996) "Search models and duration data" in *Handbook of Applied Econometrics. Vol II: Microeconomics*: 300-51. Ed. Hashem Pesaran & Peter Schmidt.

Viscusi, W.K. (1995), "Cigarette taxation and the social consequences of smoking", en J.M.Poterba ed.; *Tax Policy and the Economy* (MIT press, Cambridge, MA): 51-101.