Tema 2. Modelos con variables dependientes limitadas

1. ¿Por qué complicarnos la vida? Mínimos Cuadrados frente a Máxima Verosimilitud
   - * Maddala, capítulo 1 apartado 1.1 y 1.2
2. Modelos truncados y censurados (tipo Tobit I) y ejemplos
   - * Maddala, capítulo 6 apartados 6.1 a 6.4 y 6.9 a 6.12
   - ** Tobin, 1958
   - ** Amemiya, 1973
   - ** Amemiya, 1984
   - ** Amemiya, capítulos 9 y 10
   - ** McDonald y Moffit, 1979
   - ** Mauleón, 1987
   - ** Powell, 1986
   - ***Campa, 1995
3. Modelos de doble decisión
   - * Maddala, capítulo 9 apartados 9.1, 9.2 y 9.6 a 9.9
   - ** Amemiya, 1984
   - ** Cragg, 1971
   - ** Blundell y Meghir, 1986, 87
   - ** García y Labeaga, 1996
   - *** Huerta y Labeaga, 1992
4. Estimación en dos etapas
   - * Maddala, capítulo 8 apartados 8.1 a 8.6
   - ** Heckman, 1979
   - *** Arrazola, de Hevia y Mato, 1992
5. Modelos truncados y censurados simultáneos
   - * Maddala, capítulo 7 apartados 7.1 a 7.6

- *** Labeaga y Martínez-Ros

6. Modelos truncados y censurados y datos de panel

- ** Arellano y Bover, 1997
- ** Arellano, Bover y Labeaga, 1998
- ** Labeaga, 1993

7. Ejemplos

1. ¿Por qué complicarnos la vida? Mínimos Cuadrados frente a Máxima Verosimilitud

Este apartado lo vamos a ilustrar con un ejemplo muy sencillo para que podamos ver las consecuencias de utilizar uno u otro método. Supongamos que estamos interesados en realizar una regresión con una muestra de 10 observaciones (totalmente inventada) de las cuales el 40 por ciento son ceros, de la variable y en la variable x (da igual si ponemos constante o no ya que es una ilustración). Podemos realizar un sencillo análisis mediante Mínimos Cuadrados Ordinarios (utilizando todas las observaciones) o realizando una regresión por MCO (utilizando sólo las observaciones positivas). Los resultados de estas regresiones los podemos observar en los siguientes dos gráficos.

Gráfico 1. MCO con todas las observaciones

Gráfico 2. MCO con las observaciones positivas

En términos teóricos, comparando el estimador de MV y el estimador de MC podemos recoger en las siguientes expresiones el **sesgo** que se comete al estimar una relación no lineal como la anterior por MC:

$$y_i^* = b' X_i + u_i \qquad [1]$$

Si ahora estuviéramos interesados en nuevas variables **dependientes**, es decir, calcular el número y la proporción de empresas innovadoras por tramos de tamaño, sector y por si tienen problemas de financiación o no, podríamos construir tablas cruzadas o estadísticos descriptivos con las tres variables (o más variables). Aquí es dónde tienen sentido los modelos de elección discreta. Por ejemplo, supongamos que queremos calcular la probabilidad de innovar (no condicionada) de las empresas industriales españolas). Aunque nos adelantemos a los modelos que se describen más tarde, adelantamos que podríamos estimar un modelo **logit** (poniendo sólo una constante entre las variables explicativas). El modelo logit se podría expresar como:

$$y_i^* = b' X_i + u_i \qquad [1]$$

donde y_i^* es una variable latente (no observada) que se relaciona con la variable observada (y_i) de la siguiente forma:

$$\begin{aligned} y_i &= 1 \quad si \quad y_i^* > 0 \\ y_i &= 0 \quad en \quad caso \quad contrario \end{aligned} \qquad [2]$$

de forma que la probabilidad de que se produzca el suceso que nos interesa (probabilidad de innovar) se puede calcular como:

$$\Pr(y_i = 1) = F(b' X_i) \qquad [3]$$

y dada la distribución asociada al logit (**distribución logística**):

$$F(b' X_i) = \frac{\exp(b' X_i)}{1 + \exp(b' X_i)} \qquad [4]$$

2. Modelos truncados y censurados (tipo Tobit I) y ejemplos

Entenderemos a lo largo de todo el seminario que un modelo **truncado** es aquel para el que no se observan los valores de ninguna variable a partir de cierto umbral. Un modelo **censurado** es aquel para el que se observa el valor de todas las variables explicativas para todas las observaciones pero sólo se observa el valor de la variable de interés (dependiente) para un rango de ellas.

Ejemplos de modelos truncados:

El PSID (Panel Study of Income Dynamics) recogió en 1968 información sobre las características de un grupo de hogares (aquellos cuya renta era inferior a cierto valor) para realizar un experimento de simulación acerca de incluir tipos negativos en el impuesto sobre la renta en EE.UU. Si quisiéramos utilizar estos datos para estimar una ecuación de ingresos como la siguiente:

$$y_i = f(educación, edad, \exp eriencia, etc.) \qquad [1]$$

debiéramos tener en cuenta que la información tanto para la variable dependiente (ingresos) como para las explicativas (educación, edad, experiencia, etc.) está truncada por el valor a partir del cual no se recogen ingresos, tal como se observa en el siguiente gráfico:

Gráfico 3. Distribución normal truncada

Supongamos que estuviéramos interesados en ajustar una ecuación de ingresos en función únicamente de la educación (X_i) tal como la siguiente:

$$y_i = bX_i + u_i \qquad [2]$$

Supongamos además que u_i se distribuye de acuerdo a una distribución normal y que el punto de truncamiento es constante (k). Observamos los ingresos si $y_i < k$. Esta condición implica que $bX_i + u_i < k$ o bien que $u_i < k - bX_i$. De esta forma (tal como se observa en el gráfico 3), $E(u_i / u_i < k - bX_i)$ no es igual a cero. Además, como más adelante veremos esta media depende de las X's y consecuentemente está correlacionada con ellas por lo que la estimación por MCO de b produce estimadores inconsistentes.

Ejemplos de modelos censurados:

1. Supongamos que queremos analizar los determinantes del número de horas trabajadas por las mujeres españolas y que tenemos una muestra de la Encuesta de Población Activa (EPA). Estamos interesados en la siguiente ecuación:

$$y_i^* = b' X_i + u_i \qquad [3]$$

donde la diferencia con la ecuación anterior es que la variable de interés no es observada. Esto es así porque en la muestra de la EPA observamos un porcentaje (importante) de mujeres que declaran trabajar cero horas, como consecuencia de que la utilidad de trabajar es inferior a la utilidad de no trabajar, o dicho de otra forma, el salario de reserva es superior al salario de mercado (donde en el salario de reserva están los costes de tener alguien que cuide los niños, de desplazarse al trabajo, etc.).[1]

2. Supongamos que queremos realizar una regresión entre la cantidad exportada por las empresas industriales españolas y los posibles determinantes. De nuevo estamos como en el caso anterior, que una parte (generalmente no desdeñable) de las empresas, declara que su cantidad exportada es cero. Es necesario, tener en cuenta el comportamiento de las mismas ya que, por ejemplo, si estas empresas que no exportan exportaran, los precios de venta de los productos podrían ser menores y esto afectar el comportamiento de las que exportan (es decir, no se puede libremente eliminar dichas observaciones tal como ha quedado reflejado en la sección anterior).

3. Gastos dedicados a I+D

4. Demanda de tabaco.

Es evidente que antes de proceder al análisis de la ecuación en la que estamos interesados, hemos de conocer el verdadero carácter de las observaciones cero que las encuestas proporcionan. Pensemos en varios ejemplos. **Ejemplo 1**. Gastos en vestido y calzado proporcionado por las encuestas (más del 30 por ciento de observaciones que son ceros). ¿Quiere decir que existe un 30 por ciento de hogares españoles cuyos miembros van desnudos por la calle? **No**, estos ceros están generados por la infrecuencia con la que los individuos efectúan sus compras. Este no es un problema de decisión sobre compra sino un problema de errores en variables y, por tanto, en este tipo de modelos no estamos genuinamente interesados en este seminario. **Ejemplo 2**. Exportaciones. Probablemente las empresas realizan un análisis coste – beneficio antes de establecerse en mercados exteriores (crear sucursales u otras vías de comercialización de sus productos). Es posible que dentro de la muestra haya empresas genuinamente no exportadoras (deciden vender su producción únicamente en el interior, por cualquier razón, incluso después de haber realizado el análisis coste - beneficio). Sin embargo, también pudiera haber empresas que habiendo decidido exportar, vendan cero en los mercados

---

[1]Más adelante distinguiremos entre las observaciones nulas en horas trabajadas ya que posiblemente una parte de las mujeres que no trabajan, no la hacen aun deseándolo.

exteriores (tal vez porque sus vías de comercialización sean inadecuadas o su publicidad mal enfocada o cualquier otra causa como que maximicen su beneficio vendiendo cero en lugar de vender una cantidad positiva, por haber de rebajar demasiado el precio de sus productos, por ejemplo). Los primeros ceros a los que he aludido serían ceros de **no - participación** (es decir, para ningún nivel de sus variables explicativas, dicha empresa exporta porque es genuinamente no exportadora). Los ceros citados en segundo lugar los llamaríamos ceros de **soluciones de esquina** (es decir, la empresa maximiza su beneficio exportando una cantidad cero de su producción). **Ejemplo 3**. Participación en el mercado de trabajo. Existen individuos que no participan (ceros de no participación en horas) e individuos que deseando participar no lo hacen (ceros de soluciones de esquina). **Ejemplo 4**. Consumo de tabaco. Existen no fumadores (no - participación) y fumadores potenciales que no fuman (soluciones de esquina).

Es importante matizar estas cuestiones con anterioridad al planteamiento de los modelos econométricos porque el tratamiento de todos estos ceros es completamente distinto (en términos del modelo aplicable a cada caso). Así, si tenemos ceros debidos a una sola causa tendremos los modelos censurados sencillos (este apartado) y si los ceros son debidos a varias causas tendremos los modelos de doble decisión (apartado siguiente).

La distinción entre los modelos truncados y censurados la vamos a observar de forma sencilla cuando escribamos la función de verosimilitud de ambos modelos. Sin embargo, en la literatura econométrica normalmente se les llama de la misma forma (modelos truncados a ambos) porque están basados en las propiedades de una distribución que es la normal truncada. En todo este seminario, nosotros haremos referencia fundamentalmente a los modelos censurados ya que es el caso más común al que nos vamos a enfrentar cuando utilicemos datos correspondientes a encuestas a empresas o individuos.

Supongamos que estamos interesados en una ecuación como la descrita anteriormente para los modelos censurados:

$$y_i^* = b' X_i + u_i \qquad [1]$$

donde para tomar el ejemplo que desarrollaremos a lo largo de toda la sección pensemos que y^*_i es la proporción de sus ventas que una empresa dedica a exportar. Los determinantes de esta **propensión exportadora** están recogidos en X_i (variables ficticias sectoriales, variables ficticias de tamaño, concentración del mercado, variable ficticia de realización de innovaciones, etc.). Como en el seminario anterior, de una variable de tramos de tamaño **(tram)** que proporciona la información de la Encuesta sobre Estrategias Empresariales (ESEE):

1. < 10 trabajadores
2. de 10 a 50
3. de 50 a 100
4. de 100 a 200
5. de 200 a 500
6. > de 500 trabajadores

definamos seis variables ficticias (**t1**, **t2**, **t3**, **t4**, **t5** y **t6**) correspondientes respectivamente a los seis tramos anteriores. Asimismo, de la variable sectorial (**sector**) proporcionada por las empresas de la ESEE, definamos las siguientes cinco variables ficticias sectoriales:

1. Empresas químicas
2. Empresas eléctricas
3. Maquinaria y elementos de precisión
4. Alimentos, bebidas e industrias textiles
5. Piel, cuero y calzado

con **s1** = 1 si la empresa pertenece al sector de las empresas químicas y 0 en caso contrario y así sucesivamente para **s2**, **s3**, **s4** y **s5** con las empresas de los sectores 2, 3, 4 y 5 anteriores, respectivamente.

Presentamos en los resultados siguientes tres modelos estimados para la propensión exportadora:

Descriptivos de la variable ficticia de exportación

```
. sum dexp

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+-----------------------------------------------------
    dexp |     923    .3055255   .4608794          0          1
```

mco con todas las observaciones

```
. regress wexp t1 t2 t3 t5 t6 s1 s2 s4 s5 dinn cr4

  Source |       SS       df       MS                  Number of obs =     923
---------+------------------------------               F( 11,   911) =    1.02
   Model |  .230715916    11  .020974174               Prob > F      =  0.4288
Residual |  18.7868626   911  .020622242               R-squared     =  0.0121
---------+------------------------------               Adj R-squared =  0.0002
   Total |  19.0175785   922  .020626441               Root MSE      =   .1436

------------------------------------------------------------------------------
    wexp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      t1 |   -.012945   .0177492     -0.729   0.466       -.047779     .021889
      t2 |   .0057139   .0186442      0.306   0.759      -.0308767    .0423045
      t3 |    .001719   .0229206      0.075   0.940      -.0432644    .0467024
      t5 |  -.0164581   .0186352     -0.883   0.377       -.053031    .0201147
      t6 |   .0109585    .021161      0.518   0.605      -.0305714    .0524884
      s1 |  -.0216253   .0174719     -1.238   0.216      -.0559152    .0126646
      s2 |  -.0304097   .0216212     -1.406   0.160      -.0728429    .0120235
      s4 |   -.010053   .0179623     -0.560   0.576      -.0453053    .0251992
      s5 |  -.0166151   .0172215     -0.965   0.335      -.0504134    .0171833
    dinn |   .0118371    .009958      1.189   0.235      -.0077061    .0313803
     cr4 |  -.0781605   .0361132     -2.164   0.031      -.1490352   -.0072858
   _cons |   .1029428    .030238      3.404   0.001       .0435986     .162287
------------------------------------------------------------------------------
```

mco con todas las submuestra de observaciones positivas

```
. regress wexp t1 t2 t3 t5 t6 s1 s2 s4 s5 dinn cr4 if dexp==1

  Source |       SS       df       MS                  Number of obs =     282
---------+------------------------------               F( 11,   270) =    1.01
   Model |  .560529764    11  .050957251               Prob > F      =  0.4360
Residual |  13.5983563   270  .050364282               R-squared     =  0.0396
---------+------------------------------               Adj R-squared =  0.0005
   Total |   14.158886   281  .050387495               Root MSE      =  .22442

------------------------------------------------------------------------------
    wexp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      t1 |  -.0619346   .0511573     -1.211   0.227      -.1626525    .0387833
      t2 |   -.015783   .0540821     -0.292   0.771      -.1222592    .0906932
```

```
      t3 |   .0097836   .0692038      0.141   0.888      -.1264641    .1460313
      t5 |  -.0529631   .0553289     -0.957   0.339      -.1618941    .0559678
      t6 |   .0032604   .0613442      0.053   0.958      -.1175134    .1240342
      s1 |   -.050611   .0498749     -1.015   0.311      -.1488042    .0475822
      s2 |  -.0518371   .0630998     -0.822   0.412      -.1760673    .0723932
      s4 |  -.0252146   .0523001     -0.482   0.630      -.1281825    .0777533
      s5 |  -.0523622   .0479657     -1.092   0.276      -.1467967    .0420722
    dinn |   .0412094   .0280636      1.468   0.143      -.0140419    .0964607
     cr4 |   -.153461   .1018688     -1.506   0.133      -.3540193    .0470972
   _cons |   .2890678   .0859581      3.363   0.001       .1198345    .4583012
------------------------------------------------------------------------------
```

estimación máximo verosímil (modelo Tobit)

```
. tobit wexp t1 t2 t3 t5 t6 s1 s2 s4 s5 dinn cr4, ll(0)

Tobit Estimates                                       Number of obs =     923
                                                      chi2(11)      =    8.05
                                                      Prob > chi2   = 0.7089
Log Likelihood = -376.42883                           Pseudo R2     = 0.0106

------------------------------------------------------------------------------
    wexp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      t1 |   .0031071    .049488      0.063   0.950      -.0940165    .1002306
      t2 |    .027372   .0518617      0.528   0.598      -.0744101    .1291542
      t3 |  -.0113858   .0648138     -0.176   0.861      -.1385873    .1158158
      t5 |  -.0335146   .0526318     -0.637   0.524      -.1368081    .0697789
      t6 |   .0372099   .0587545      0.633   0.527      -.0780999    .1525197
      s1 |  -.0437939   .0482787     -0.907   0.365      -.1385441    .0509564
      s2 |  -.0741144   .0607744     -1.220   0.223      -.1933883    .0451595
      s4 |  -.0205537   .0503028     -0.409   0.683      -.1192765     .078169
      s5 |  -.0161093   .0472855     -0.341   0.733      -.1089104    .0766917
    dinn |    .017251   .0275557      0.626   0.531       -.036829    .0713311
     cr4 |  -.1821272    .100287     -1.816   0.070      -.3789474    .0146929
   _cons |  -.0782187   .0840749     -0.930   0.352      -.2432214     .086784
---------+--------------------------------------------------------------------
     _se |   .3205997    .015064            (Ancillary parameter)
------------------------------------------------------------------------------

Obs. summary:       641 left-censored observations at wexp<=0
                    282 uncensored observations
```

Nota. Ver, por ejemplo, lo que sucede con la concentración.

¿Qué tipo de modelo estamos estimando en cada caso?

$$E(y_i / X_i) = \sum y_i \Pr(y_i) = 1 \quad \Pr(y_i = 1) + 0 \quad \Pr(y_i = 0) = \qquad [7]$$

$$\Pr(y_i = 1) = b' X_i$$

y como consecuencia, cuando hayamos estimado los parámetros, la expresión

$$\hat{y}_i = \hat{b}' X_i \qquad [8]$$

nos dará la probabilidad estimada del suceso $y_i = 1$ dados unos valores concretos de X_i. Esta probabilidad, por definición deberá estar en el intervalo cerrado [0, 1]. Veámoslo:

$$\begin{array}{ccc} y_i & u_i & f(u_i) \\ 1 & 1 - b' X_i & b' X_i \\ 0 & -b' X_i & 1 - b' X_i \end{array} \qquad [9]$$

es decir, nada garantiza que la media de la variable de interés sea inferior a 1 y superior a 0 como se requiere de una probabilidad. Además,

$$\begin{aligned} & Var(u_i) = b' X_i (1 - b' X_i)^2 + (1 - b' X_i)(b' X_i)^2 = \\ & b' X_i (1 - b' X_i) = E(y_i)[1 - E(y_i)] \end{aligned} \qquad [10]$$

de nuevo nada garantiza que la varianza sea positiva (como se requiere) y, además, el modelo presenta heterocedasticidad como se aprecia en la expresión [10].

Finalmente, de la expresión [9] se deduce que los residuos no se distribuyen de acuerdo a una Normal (aunque esto no es un problema grave).

Para describir este problema realizamos una regresión lineal de la variable dependiente nipos (realiza innovaciones de proceso) en una serie de variables explicativas en la submuestra de empresas de más de 500 trabajadores (t6 = 1) con los siguientes resultados:

$$y_i^* = b' X_i + u_i \qquad [11]$$

y_i^*, como se ha dicho es no observada (no observable). En los datos observamos y_i, y la relación entre ambas es de la siguiente forma:[2]

$$\begin{aligned} y_i &= 1 \quad si \quad y_i^* > 0 \\ y_i &= 0 \quad en \quad caso \quad contrario \end{aligned} \qquad [12]$$

de forma que la probabilidad de que se produzca el suceso que nos interesa (probabilidad de innovar) se puede calcular como:

$$\Pr(y_i = 1) = F(b' X_i) \qquad [13]$$

Vemos, por tanto, que hemos establecido una diferencia fundamental entre [13] y [6]. En [6] $E(y_i / X_i) = b' X_i$ y, sin embargo en [13] $E(y_i^* / X_i) = b' X_i$. Pero esta expresión no nos interesa sino que nos interesa:

$$E(y_i / X_i) = \Pr(u_i > -b' X_i) \qquad [14]$$

$$F(u_i) = \frac{\exp(u_i)}{1 + \exp(u_i)} \qquad [15]$$

$$\Pr(y_i = 1) = \Pr(u_i > -b' X_i) = 1 - F(-b' X_i) = F(b' X_i) \qquad [16]$$

[2]Notar que esta formulación es sencilla en términos de utilidad o de beneficio. Es decir, el individuo elige la alternativa 0 si le produce mayor utilidad que la 1. Aunque la utilidad de ambas alternativas no se observa, sí se observa la decisión porque en los datos vemos la alternativa elegida. De la misma forma, la empresa elige innovar si dicha acción le produce un mayor beneficio. De hecho, dado que observamos si la empresa innova o no, observaremos el beneficio dada la acción llevada a cabo pero no en el otro caso, por lo que podemos asumir la formulación en función de variables no observadas o latentes.

$$\Pr(y_i = 0) = F(-b'X_i) = 1 - F(b'X_i) \qquad [17]$$

Ahora expresamos la probabilidad de las N observaciones muestrales (que llamaremos **función de verosimilitud**):

$$L = \prod_{i=1}^{N} \left[F(b'X_i)\right]^{y_i} \left[1 - F(b'X_i)\right]^{1-y_i} \qquad [18]$$

donde el productorio proviene de la independencia entre las observaciones. Obtener los estimadores MV del vector de parámetros b consiste en obtener los valores de los mismos que maximizan [18] o lo que es lo mismo, que satisfacen las condiciones de primer orden o gradientes.[3] Como es lo mismo optimizar una función que una transformación monótona creciente de la misma, una práctica habitual suele ser tomar logaritmos en [18] y optimizar el logaritmo de la función de verosimilitud:

$$\log L = \sum_{i=1}^{N} \left[y_i \log\left[F(b'X_i)\right] + (1-y_i)\log\left[1 - F(b'X_i)\right]\right] \qquad [19]$$

donde log hace referencia a logaritmo natural (ln).

Valores observados $\rightarrow$ **Valores predichos** $\downarrow$	**1**	**0**	
1	Acierto	Error	Nº unos predicho

[3]De esta expresión debe quedar claro que no se pueden identificar b y s por separado por lo que bien asumimos que $s = 1$ o que lo que podemos estimar es el ratio b/s. No entraremos en detalle de cómo se obtienen los estimadores. Se utilizan métodos de optimización no lineal que consisten en partir de unos valores iniciales para los parámetros y continuar mediante un procedimiento iterativo hasta que se obtiene convergencia (en los parámetros, en el valor de la función en el máximo y en los gradientes).

0	Error	Acierto	Nº ceros predicho
	Nº unos observ.	Nº ceros observ.	

Modelo de probabilidad lineal:

$$\frac{\partial E(y_i / X_i)}{\partial X_j} = b_j \qquad [20]$$

Modelo logit:

$$\frac{\partial E(y_i / X_i)}{\partial X_j} = \frac{\exp(b' X_i)}{[1 + \exp(b' X_i)]^2} b_j \qquad [21]$$

Modelo probit:

$$\frac{\partial E(y_i / X_i)}{\partial X_j} = f(b' X_i) b_j \qquad [22]$$

siendo f(.) la función de densidad de probabilidad de la normal estándar.

A la vista de las expresiones [20]-[22], el coeficiente de la variable cuyo efecto queremos medir no mide más que el efecto sobre la variable de interés en el modelo de probabilidad lineal (algo que ya sabemos por el modelo de regresión lineal), mientras que para medir el efecto verdadero se debe ponderar por una medida de la probabilidad en cada caso.

6. Modelos truncados y censurados simultáneos

Supongamos que en el modelo planteado en [1], uno de los condicionantes es una variable endógena (cualitativa o no). El tratamiento que se debiera dar al modelo sería totalmente diferente. En Labeaga y Martínez-Ros (1994) se plantea un modelo de determinación de decisión de innovar, empleo y decisión de exportar y se estima utilizando datos de la CB para

1990. Todo el tratamiento algebraico es bastante complicado por lo que nos remitimos a dicho artículo o al capítulo 5 (apartados 5.7 a 5.9) del libro de Maddala.

7. Modelos truncados y censurados y datos de panel

Terminaremos con la estimación de dos modelos de elección discreta, logit de efectos fijos y probit de efectos aleatorios, que se pueden aplicar cuando se dispone de datos de panel y se tiene interés en explicar el comportamiento de una variable discreta. El planteamiento de los modelos sería similar al utilizado en la sección tercera, pero el tratamiento econométrico es completamente diferente. Ahora al disponer de varias observaciones temporales para cada unidad muestral (empresa), se pueden controlar las variables no observables (habilidad de los dirigentes de las mismas, efectos de experiencia o cualquier variables que no tenga variación temporal y no tengamos entre la información). Esto es muy importante porque esos efectos no observables pueden estar correlacionados con las variables incluidas como explicativas en el modelo y su no control sesgará los parámetros obtenidos. El tratamiento econométrico es, por tanto, mucho más difícil y es muy extraño encontrar artículos donde simultáneamente se tengan en cuenta los problemas de variable dependiente cualitativa y se haga además en un contexto de panel. En los ejemplos que sigue se presentan los resultados de estimar una ecuación de decisión de innovar en producto y se dispone del panel de datos de la ESEE durante el período 1990-93. Se han de destacar varios aspectos de los resultados:

1. El modelo logit de efectos fijos está estimado por MV pero tras realizar una transformación que elimina los efectos. De ahí que para todas las variables que no tienen variación temporal no está identificado el parámetro (en el caso que nos ocupa las variables sectoriales y la constante). El modelo probit de efectos aleatorios se estima por MV pero en niveles, de ahí que estén identificados incluso los parámetros que acompañan a variables sin variación temporal. Lo que se supone que los efectos son parte de la perturbación y se modeliza dicho error con un esquema de autocorrelación (se supone equi-correlación).
2. En el modelo logit de efectos fijos se supone que los mismos están potencialmente correlacionados con las variables incluidas en la especificación, de ahí que se utiliza una transformación para eliminarlos. En el modelo probit de efectos aleatorios al formar parte

los efectos de la perturbación se debe suponer que variables y efectos no están correlacionados para obtener estimadores consistentes.

3. La posibilidad de eliminar los efectos para la estimación de los parámetros en el modelo logit, se da porque existe un estadístico suficiente para poder hacerlo. Este estadístico es la suma del indicador de realizar innovaciones para cada individuo durante los T períodos para los que se observa. Sabemos que si dicha suma es 4 (número de períodos de la muestra) la probabilidad de realizar innovación cada período es 1. Si la suma es 0, la probabilidad de realizar innovaciones en dicha empresa es 0. Así que los sucesos {realiza innovaciones todos los períodos de la muestra} y {no realiza innovaciones en ningún período} no se tienen en cuenta durante la estimación porque contribuyen con una probabilidad cero al logaritmo de la función de verosimilitud. Solo contribuyen los sucesos para los que existe cambio de régimen (innova – no innova). En el modelo probit de efectos aleatorios, todas las observaciones se tienen en cuenta (para comprobar ver el número de observaciones que aparece en cada uno de los listados de ordenador, además de las notas que aparecen al principio del listado del logit de efectos fijos).[4]

[4]Una descripción muy simple de ambos modelos se puede consultar en Maddala (capítulo 2, apartado 2.17).