Master en Estadística Aplicada y Estadística para el Sector Público

Modelos Lineales

Tomo 2

Ramón Mahía Cansado
Juan Muro Romero

Universidad de Alcalá
Instituto Nacional de Estadística
CIFF
Centro Internacional de Formación Financiera
2008
2009

<u>INTRODUCCIÓN A LA ESPECIFICACIÓN Y ESTIMACIÓN DE MODELOS CON DATOS DE PANEL</u>

Ramón Mahía
Dpto. ECONOMÍA APLICADA
UNIVERSIDAD AUTÓNOMA DE MADRID
ramon.mahia@uam.es
www.uam.es/ramon.mahia

1.- CUESTIONES GENERALES

1.A.- DEFINICIÓNES EN TORNO A LOS MODELOS DE DATOS DE PANEL

- **La técnica se encuadra** en el análisis de regresión, incluida en el conjunto de herramientas multivariantes destinadas al análisis de la dependencia entre variables, medidas todas ellas (endógena y exógenas) preferentemente en una escala estrictamente cuantitativa.

- El término **modelo de datos de panel** se aplica en este contexto a aquel modelo de regresión que utiliza, para la estimación de los parámetros de interés, la variabilidad temporal y transversal de los datos.

- **Ejemplos** de modelos con datos de panel podrían ser:

  1. Un análisis de la influencia de determinadas variables (renta, tamaño familiar, composición familiar) en el consumo alimenticio en las familias españolas utilizando los datos de la Encuesta de Presupuestos Familiares correspondientes a diferentes momentos del tiempo.

  2. Un análisis del cumplimiento de la hipótesis de la paridad del poder adquisitivo en los países de la OCDE tomando los datos de precios y tipos de cambio nominales correspondientes a los últimos 40 años.

- Generalmente**, los paneles de datos se distinguen unos de otros según su amplitud transversal y profundidad temporal**. Así, los paneles con un número muy amplio de observaciones transversales (ejemplo 1 anterior) se denominan Paneles Micro, mientras que los paneles centrados en una amplia dimensión temporal se suelen denominar Paneles Macro. En el caso, realmente extraordinario, de contar con un panel con amplia dimensión tanto temporal como transversal hablaríamos de un "Campo Aleatorio" o Random Field.

- Así mismo, resulta habitual hablar de **paneles de datos equilibrados** cuando el número de observaciones transversales es el mismo para cada período del tiempo y de **paneles completos** cuando el número de observaciones temporales es el mismo para cada elemento del panel.

- Es importante dejar claro que, en sentido estricto, **no son datos de panel** los paneles rotatorios o la mera agregación de cortes transversales independientes[1]. Para construir un elemento verdaderamente útil de cara a la inferencia, se trata de que la variabilidad temporal y transversal corresponda a una misma muestra de individuos para todas las observaciones. En este sentido, en el caso del análisis empresarial, la gran heterogeneidad dificulta la construcción de verdaderos paneles.

1.B.- QUÉ JUSTIFICA EL INTERÉS POR LOS DATOS DE PANEL

- **La utilización de datos de panel en lugar de series temporales se justifica por aprovechar la variabilidad transversal.** La identificación y estimación de los parámetros de una función de respuesta explota la variación de las variables incluidas. Si las variables no presenta presentan excesiva variabilidad temporal pero sí transversal, la aproximación con datos de panel aportaría capacidad extra para esa estimación. *Por ejemplo, si estamos interesados en describir el comportamiento del flujo internacional de turistas de unas zonas a otras, seguramente encontraremos variables con suficiente variación temporal como los movimientos de los tipos de cambio bilaterales o la renta de los turistas, sin embargo, la influencia de variables como el clima, no podrá observarse si no es comparando unos países con otros.*

- **La utilización de datos de panel en lugar de series transversales se justifica por aprovechar la variabilidad temporal para:**

---

[1] A este tipo de conjuntos de datos se los denomina pseudo - paneles existiendo algunas formas de aprovechar eficientemente su información entre las que destacan las propuestas por Deaton (1985).

1.- De modo simétrico a lo expuesto en el punto anterior, algunas variables pueden presentar variabilidad temporal pero no transversal de modo que su efecto sólo podría captarse con dimensión temporal.

2.- Si disponemos de más de un corte temporal para **los mismos individuos**, la búsqueda de la eficiencia sugeriría una estimación con todo el panel de datos. En este sentido, resulta práctico tener en mente la siguiente relación sencilla:

$$V(\beta) = \frac{V(U)}{V(X)}$$

3.- Por otro lado, disponer de más de una observación temporal permitirá controlar (separar) la heterogeneidad transversal **inobservable** utilizando transformaciones apoyadas precisamente en disponer de más de una observación temporal (*diferencias simples, diferencias ortogonales* ...)

- En algunos casos específicos, la introducción de la variabilidad transversal en procedimientos clásicos de análisis meramente temporal, favorece las propiedades estadísticas de los procedimientos de inferencia tradicionales. Así, por ejemplo, la potencia y propiedades asintóticas de los test de integración y cointegración se ven claramente favorecidas cuando se combinan datos de corte temporal y transversal, lo que justifica el actual interés por las líneas metodológicas que combinan análisis de series temporales y utilización de variabilidad transversal.

1.C.- PRESENTACIÓN MATEMÁTICO–ESTADÍSTICA DE UN MODELO CON DATOS DE PANEL

- **Expresión general**

Un modelo de datos de panel podría formularse en términos plenamente genéricos como:

$$y_{it} = \eta_{it} + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + ... + \beta_{ki} x_{kit} + v_{it}$$

i=1,2....N
t=1,2..........T

permitiendo la presencia de múltiples parámetros individuales y temporales y una definición sin restricciones sobre la composición y propiedades del vector de perturbaciones aleatorias . Sin embargo, tal representación no resulta viable por lo que, generalmente, se recurre a formulaciones más restrictivas tanto en términos paramétricos como con relación a los supuestos sobre el vector de observaciones. Siguiendo a Johnston (1992) podemos ordenar 7 tipos de especificaciones en una “taxonomía” sobre los modelos de datos de panel:

	Supuestos sobre:		
Modelo	**Ordenada en el origen “η”**	**Coeficientes de pendiente “β”**	**Vector de perturbaciones “v”**
I(a)	Común en “i” y “t”	Común en “i” y “t”	Matriz de varianzas y covarianzas escalar
I(b)	Común en “i” y “t”	Común en “i” y “t”	Matriz de varianzas y covarianzas NO escalar
II(a)	Variando en “i”	Común en “i” y “t”	**Efectos fijos**
II(b)	Variando en “i”	Común en “i” y “t”	**Efectos aleatorios**
III(a)	Variando en “i” y “t”	Común en “i” y “t”	Efectos fijos
III(b)	Variando en “i” y “t”	Común en “i” y “t”	Efectos aleatorios
IV	Variando en “i”	Variando en “t”	Matriz de varianzas y covarianzas escalar o NO

- **Modelo de “efectos fijos” y modelo de “efectos aleatorios”**

De entre todos los modelos presentados en la tabla anterior, dos de ellos son los más utilizados tradicionalmente: los modelos II(a) y II(b). El modelo II(a) se denomina “Modelo de efectos fijos” y el modelo II(b) “Modelo de efectos aleatorios” haciendo referencia al modo en que se considera la heterogeneidad inobservable transversal en su relación con la perturbación aleatoria.

Contrariamente a lo que podría parecer tras una lectura rápida de un buen número de textos econométricos, de lo anterior debe deducirse que la diferencia entre efectos fijos o aleatorios no radica en la morfología del modelo, que es siempre la siguiente:

$$y_{it} = \beta x_{it} + \eta_i + v_{it}$$

donde η_i representa la heterogeneidad transversal inobservable, sino que lo trascendental radica en si esa heterogeneidad se considera fija, determinista, o por el contrario se define como la composición de una parte fija común, más una aleatoria específica para cada individuo.

2.- ESTIMACIÓN CON DATOS DE PANEL ESTÁTICOS

2.A.- ESTIMACIÓN DEL MODELO II(A) DE EFECTOS FIJOS

Suponemos en este caso que existe heterogeneidad transversal inobservable, constante en el tiempo de carácter no aleatorio y permitiremos la presencia de términos independientes diferentes η_i para cada individuo del panel. En este contexto, la estrategia para la estimación de parámetros podría ser:

1.- Utilizar el estimador MCO tradicional sobre el modelo en niveles utilizando variables ficticias de grupo. El problema es la gran cantidad de parámetros a estimar dada la gran cantidad de variables ficticias a utilizar (N-1). Una de las ventajas de esta estrategia de estimación radica en que se obtienen de forma direca intervalos de confianza para los efectos fijos estimados.

2.- Transformar el modelo en diferencias temporales clásicas y aplicar MCO. El defecto de este método es que la transformación en diferencias provocaría la aparición de autocorrelación en los residuos de las ecuaciones en diferencias (**salvo en el caso T=2**):

$$\left.\begin{aligned} \Delta v_2 &= v_2 - v_1 \\ \Delta v_3 &= v_3 - v_2 \end{aligned}\right\} \rightarrow Cov(\Delta v_t \Delta v_{t-1}) = -\sigma^2 \quad y \quad V(\Delta v_t) = 2\sigma^2$$

lo que podría resolverse recurriendo a una estimación MCG calculando previamente la nueva matriz no escalar de varianzas y covarianzas Ω para “v” (suponiendo que no existía previamente autocorrelación en “v”):

$$\hat{\beta} = \frac{\sum_i \Delta X_i' \Omega^{-1} \Delta Y_i}{\sum_i \Delta X_i' \Omega^{-1} \Delta X_i} \quad \text{con} \quad \Omega = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{pmatrix}$$

3.- Utilizar MCO sobre el modelo en desviaciones ortogonales. Las desviaciones ortogonales implican una transformación sobre los datos X e Y equivalente a la ponderación Ω del producto X’Y del procedimiento anterior.

4.- Utilizar el estimador **MCO sobre el modelo en diferencias respecto a las medias grupales,**

$$\left(y_{it} - \bar{y}_i\right) = \beta(x_{it} - \bar{x}_i) + \left(v_{it} - \bar{v}_i\right)$$

esta transformación permitiría concentrar la estimación en un único conjunto de parámetros “β” pudiendo calcularse posteriormente los parámetros η_i individuales simplemente como:

$$\hat{\eta}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{1i} - \hat{\beta}_2 \bar{x}_{2i} - \ldots - \hat{\beta}_k \bar{x}_{ki}$$

Esta estrategia resulta generalmente válida por lo que es en la práctica la forma más común de resolver la estimación de un modelo de efectos fijos. El estimador anterior, recibe el nombre de **Estimador Intragrupos** por considerarse para su aplicación la desviación intragrupal para X e Y. También recibe el nombre de estimador de efectos fijos o de covarianzas.

2.B.- ESTIMACIÓN DEL MODELO II(B) DE EFECTOS ALEATORIOS

En el modelo de efectos aleatorios se supone una sola ordenada en el origen "α" y las N ordenadas específicas correspondientes a cada individuo del panel se integran en la perturbación aleatoria de modo que:

$$v_{it} = \eta_i + \varepsilon_{it}$$

por esta razón a estos modelos se les llama también modelos con errores compuestos (*error components models*).

La estimación de los parámetros de un modelo de efectos aleatorios no requiere, como en el caso de efectos fijos, diferencias o desviaciones o cualquier otra transformación que elimine la presencia del efecto fijo "η_i". Muy al contrario, junto con la estimación de los parámetros "α" y "β" interesa la estimación diferenciada de la varianza de los efectos aleatorios σ_η dentro de la estimación global de la varianza de la perturbación aleatoria, por lo que no conviene que el efecto "η_i" sea obviado en el procedimiento de estimación. Llegados a este punto, podría pensarse en utilizar el estimador simple MCO sobre el modelo en niveles:

$$y_{it} = \alpha + \beta x_{it} + (\eta_i + \varepsilon_{it})$$

Sin embargo, esta estrategia de estimación no resulta válida ya que la presencia de un efecto temporalmente constante en la perturbación aleatoria "η_i" provoca autocorreación residual y modifica la expresión tradicional de la varianza de la perturbación aleatoria. Así, la matriz de varianzas y covarianzas de la perturbación queda:

$$\Omega = \begin{pmatrix} \sigma_\eta^2 + \sigma_\varepsilon^2 & \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\varepsilon^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \ddots & \vdots \\ \sigma_\eta^2 & \sigma_\eta^2 & \cdots & \sigma_\eta^2 + \sigma_\varepsilon^2 \end{pmatrix} = \sigma_v^2 \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

con:

$$\rho = \frac{\sigma_\eta^2}{\sigma_v^2}$$

Por tanto, la estrategia correcta consiste en **usar MCG sobre el modelo en niveles utilizando** una estimación adecuada de la anterior matriz Ω. A este estimador se le denomina estimador **Balestra-Nerlove ó Estimador Entre Grupos,** nombre este último que proviene de la equivalencia entre estos resultados y los que se obtendrían planteando la estimación MCO entre las medias grupales de "y" y "x". Como siempre, debemos recordar que existe una transformación previa sobre los datos que permite su correcta estimación simple por MCO evitando la utilización del estimador MCG (*aunque nunca eludiendo la estimación de las varianzas σ^2_η y σ^2_v*).

La estimación Balestra - Nerlove requerirá por tanto necesariamente dos etapas:

1.- Estimación de un modelo previo del que utilizar los residuos para aproximar σ^2_η y σ^2_v. Normalmente suele utilizarse, o bien una estimación MCO simple del modelo, con término independiente "α" común, o bien una estimación del modelo con efectos fijos.

2.- **Estimación de las varianzas σ^2_η y σ^2_v y aplicación del estimador MCG**. Para este paso, partiendo de:

$$v_{it} = \eta_i + \varepsilon_{it}$$

utilizaremos una aproximación por descomposición tradicional de la varianza del siguiente modo:

$$\sum_{i,t}\left(v_{it} - \overline{\overline{v}}\right)^2 = \sum_{i,t}\left(v_{it} - \overline{v}_i\right)^2 + \sum_{i,t}\left(\overline{v}_i - \overline{\overline{v}}\right)^2$$

Para el primer término, la media cuadrática intragrupal, tenemos que:

$$\sum_{i,t}\left(v_{it} - \overline{v}_i\right)^2 = \sum_{i,t}\left(\varepsilon_{it} - \overline{\varepsilon}_i\right)^2$$

Utilizando ahora el concepto de cuasivarianza muestral que, supuesta la distribución $N(0,\sigma^2_x)$ para una muestra x_1, x_2,....x_T nos dice que:

$$E\left[\frac{\sum\left(x_{it} - \overline{x}_i\right)^2}{T-1}\right] = \sigma_x^2$$

y aplicándolo a la perturbación ε, (*supuesta la normalidad de la misma*), tenemos que:

$$E\left[\sum_{i,t}\left(\varepsilon_{it} - \overline{\varepsilon}_i\right)^2\right] = N(T-1)\sigma_\varepsilon^2$$

Para el segundo término, de modo similar, escribiremos que::

$$\sum_{i,t}\left(\overline{v}_i - \overline{\overline{v}}\right)^2 = \sum_{i,t}\left(\eta_i - \overline{\eta}\right)^2 + \sum_{i,t}\left(\overline{\varepsilon}_i - \overline{\overline{\varepsilon}}\right)^2 + 2\sum_{i,t}\left(\overline{\eta}_i - \overline{\eta}\right)\left(\overline{\varepsilon}_i - \overline{\overline{\varepsilon}}\right)$$

teniendo ahora:

$$E\left[\sum_{i=1}^{N}\left(\eta_i - \overline{\eta}\right)^2\right] = (N-1)\sigma_\eta^2$$

$$E\left[\sum_{i=1}^{N}\left(\overline{\varepsilon}_i - \overline{\overline{\varepsilon}}\right)^2\right] = (N-1)\frac{\sigma_\varepsilon^2}{T}$$ [2]

de modo que:

$$\sum_{i,t}\left(\overline{v}_i - \overline{\overline{v}}\right)^2 = (N-1)T\sigma_\eta^2 (N-1)\sigma_\varepsilon^2$$

En términos operativos el cálculo no resultaría complejo en cuanto que implicaría tan sólo el cálculo inicial de σ^2_ε:

---

[2] *En la segunda de las expresiones se entiende que cada una de las ε_i que entran en la media se han extraído de una población $N(0/T,\sigma^2_\varepsilon/T)$*

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{i,t}\left(\hat{v}_{it} - \bar{\hat{v}}_{i}\right)^{2}}{N(T-1)}$$

y a continuación el de σ^2_η como:

$$\hat{\sigma}_{\eta}^{2} = \frac{1}{T}\left(\frac{T}{N-1}\sum_{i=1}^{N}\left(\bar{\hat{v}}_{i} - \bar{\hat{v}}\right)^{2} - \hat{\sigma}_{\varepsilon}^{2}\right)$$

2.C.- LA ELECCIÓN ENTRE EFECTOS FIJOS Y EFECTOS ALEATORIOS

Frente a esta pregunta, lo más importante es recordar que la elección entre uno u otro modelo no puede tomarse basándonos exclusivamente en el resultado de un contraste estadístico de especificación sino que debe apoyarse en las peculiaridad de la realidad del fenómeno analizado. Como señalan Arellano y Bover (1990): *“....Una costumbre muy extendida en el trabajo aplicado consiste en estimar ambos modelos para a continuación contrastar si los efectos son fijos o aleatorios. Este es quizá el malentendido mas extendido en este campo.....”*

Para valorar la adecuación de uno u otro modelo a la realidad “institucional” del escenario analizado deben considerarse, no obstante, una serie de cuestiones de orden técnico que pueden pasar desapercibidas.

1.- El problema de la correlación entre la heterogeneidad inobservable η_i y los regresores x_{it}. En la mayor parte de las ocasiones, la importancia de un modelo de efectos fijos vendrá determinada, no tanto por la variabilidad transversal, por la existencia de efectos diferenciales η_i constantes en el tiempo, sino por su correlación con los regresores x_{it}. En este sentido, los efectos individuales se pueden considerar siempre aleatorio sin pérdida de generalidad.

Efectivamente, si partimos de un modelo pensado para un único corte transversal:

$$y_i = \alpha + \beta x_i + (\eta_i + \varepsilon_i)$$

y existiese correlación entre x_{it} y v_{it} , la estimación por aproximación tradicional de regresión permitiría obtener un β que sería el predictor óptimo β pero no el parámetro de simulación de efectos, es decir, el verdadero valor del parámetro que determina la respuesta de “y” ante variaciones de las “x”. Efectivamente, conviene recordar aquí que la **identificación** de los parámetros de una función de respuesta en el modelo de regresión **se apoya en las propiedades asumidas para la perturbación “u”** con relación a los regresores:

$$\begin{aligned}
&Cov_p(u,x) = 0\\
&E[U \mid X] = 0 \rightarrow Cov_p(u, f(x)) = 0\\
&X \perp U = 0 \rightarrow Cov_p(f(u), f(x)) = 0
\end{aligned}$$

en la medida en que estas propiedades sean razonables, los parámetros representarán realmente la dependencia de “y” sobre “x”:

$$\begin{aligned}
&Cov_p(u,x) = 0 \rightarrow Cov(y-\alpha-\beta x, x) = 0\\
&\rightarrow Cov(y,x) - \beta V(x) = 0 \rightarrow \beta = \frac{Cov(x,y)}{V(x)}
\end{aligned}$$

y, suponiendo además E[u]=0:

$$E[u] = 0 \rightarrow E[y-\alpha-\beta x] = 0 \rightarrow \alpha = E[y] - \beta E[x] = E[y] - \frac{Cov(x,y)}{V(x)} E[x]$$

En un contexto puramente transversal o puramente temporal podríamos optar por un enfoque de **variables instrumentales** en el que se sustituye:

$$Cov_p(u,x) \neq 0 \rightarrow Cov_p(z,u) = 0$$

siendo: $Cov_p(z,x) \neq 0$. Sin embargo, una solución alternativa consistiría en acudir a los datos de panel de modo que, al combinar la dimensión temporal y transversal, pudiésemos neutralizar el efecto de la correlación (x,u). Esto es precisamente lo que se logra cuando, suponiendo que la correlación entre x y η_i es constante en el tiempo, utilizamos el estimador Intra – Grupos en el modelo de efectos fijos utilizando el panel para separar la variación transversal permanente. El único requisito adicional consiste en que los regresores X presenten variación temporal.

Ejemplo – Caso Práctico (1): Supongamos que disponemos de un corte transversal y queremos estudiar la relación entre el logaritmo de los ingresos, el nivel de educación y la habilidad del trabajador. Suponiendo inobservable la habilidad, ¿aportaría alguna ventaja la incorporación de la dimensión temporal?. Griliches. 1977.

Solución: NO. Ya que el nivel de educación no presenta variación transversal lo que impedirá incorporar el efecto de la habilidad mediante la aplicación de estimadores IG o EG.

Ejemplo – Caso Práctico (2): Supongamos que disponemos de un corte transversal y queremos relacionar las horas trabajadas con el salario percibido. Reconociendo la importancia y el carácter inobservable de las percepciones acerca del flujo de salarios futuros sobre las horas trabajadas, ¿aportaría alguna ventaja la incorporación de la dimensión temporal?. MaCurdy, 1981.

Solución: SI. En primer lugar, dado que en este caso las variables pueden presentar suficiente variación temporal, cabe pensar en incorporar esta dimensión. El problema consiste en que, dada la más que probable relación entre salario percibido y expectativas futuras de salario, deberemos utilizar un estimador IG por lo que perdemos la posibilidad de estimar el efecto de esta variable de expectativas sobre las horas trabajadas.

2.- Si el estimador IG permite controlar el efecto de la heterogeneidad inobservable sobre los parámetros "β", ¿qué puede motivar el uso de un estimador tipo Balestra Nerlove?.

Imaginemos un corte transversal con heterogeneidad "η_i". Una estimación en ese corte transversal ya valdría para estimar "β" adecuadamente salvo que sospecháramos que "η_i" y "x_i" están correlacionadas en cuyo caso utilizaríamos un enfoque de VI. Imaginemos ahora que interesa añadir la dimensión temporal por cualquiera de las razones previamente apuntadas y que podemos hacerlo dado que x_i presenta suficiente variación temporal. En ese caso, para asegurar las posibles distorsiones de una hipotética relación "η_i" y "x_{it}" estimaríamos siempre por IG eligiendo la opción más robusta: es decir, en el caso en que realmente no existiese relación "η_i" "x_{it}", haber usado IG nos habría supuesto una pequeña pérdida de eficiencia pero en cualquier caso tendríamos siempre un estimador consistente. ¿Qué puede movernos a utilizar un estimador BN que no eliminará la presencia de "η_i"?

2.a.- La primera parte de la respuesta es obvia: que interesase considerar la influencia de ese aspecto transversal η_i con carácter temporal permanente, sobre la variable endógena, estando seguros de que la correlación entre η_i y x_{it} no existiese.

Ejemplo – Caso Práctico (3): Supongamos que disponemos de un corte transversal y queremos relacionar el input de una explotación agrícola con el output de la misma. En la perturbación aleatoria quedarían "shocks" incontrolables como el régimen de lluvias u otros cambios climáticos. Sin embargo, es posible que en la perturbación quedase incluido también un factor trascendental: la calidad del suelo. Si esa calidad del suelo fuese cuantificable se incorporaría como una variable más pero, si fuese inobservable, podríamos controlar su efecto incorporando la dimensión temporal ya que, ese efecto sería constante en el tiempo. Pero además, suponiendo la ausencia de correlación entre calidad e input, podríamos usar un estimador EG que permitiese diferenciar los efectos de la variable de CALIDAD DEL SUELO, efectos inobservables de naturaleza permanente, de otros "shocks" también de carácter inobservable, pero cuyo efecto se diluye en el tiempo. Mundlak, (1978) En este sentido, y en un plano más formal, conviene observar que al permitir la presencia de η_i en v_{it} tenemos:

$$Cov(y_{it}, y_{it-1}) = \sigma_\eta^2$$

$$Corr(y_{it}, y_{it-1}) = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2}$$

de modo **que resultaría posible distinguir de la dinámica de y_t debida a aspectos inobservables, la parte intertemporal** (permanente) **de la transitoria.**

2.b.- Además de la anterior consideración, y tal y como se apuntaba en la introducción a este punto (2) debe tener en cuenta que existe además una razón de índole estadística. Efectivamente, en caso de que el efecto η_i esté incorrelacionado con "x", el estimador BN es mejor alternativa que el IG dado que es el estimador eficiente. Sin embargo, si existe correlación entre "ηi" y "x", este estimador BN es inconsistente para "β". Por el contrario, el estimador IG es consistente en ambos casos (más robusto) aunque en el caso en que no exista correlación entre "η_i" y "x_{ij}", es menos eficiente que BN.

2.D.- USO DEL CONTRASTE DE ESPECIFICACIÓN DE HAUSMAN

Una vez admitido que la decisión EF Vs. EA debe ser responsabilidad del modelizador, ¿cuándo cederemos esa responsabilidad a un contraste estadístico como el test de Hausman?. En principio el contraste de se aplicará sobre todo en el caso de no estar seguros de la relación entre "η" y "x" el contraste de HAUSMAN sería pertinente.

En este sentido, el test de Hausman es un contraste clásico de robustez frente a eficiencia. Este tipo de contrastes plantean siempre dos estimadores para un mismo conjunto de parámetros, uno robusto θ_R consistente tanto en la hipótesis nula H_0 como en la alternativa H_1 (*cualesquiera que sean*) y otro eficiente θ_E pero sólo bajo la hipótesis nula H_0. Si, una vez calculados ambos, la diferencia observada entre los dos estimadores (θ_R -θ_E) es escasa, se toma evidencia a favor de la hipótesis nula

En nuestro contexto los estimadores son θ_R=Intra-Grupos y θ_E=Balestra –Nerlove y la hipótesis nula será que el efecto η_i esté incorrelacionado con x_{it} (H_0). Una vez estimados, **para analizar si existe o no correlación entre "η" y "x" puede verse si hay diferencias significativas entre IG y BN**. Si estamos en H_0 (ausencia de correlación) BN será consistente y IG también (*aunque algo menos eficiente*) por lo que su parecido será significativo. Si estamos en H_1, BN no será consistente por lo que su valor podrá distar del valor de IG.

Formalmente, el test de Hausman se presenta como una χ^2 formada por la diferencia relativa entre las dos estimaciones alternativas relativizadas por la varianza de esa diferencia. Valores de la χ^2 superiores a la referencia de tabla indican la presencia de correlación entre η_i y x_{it}.

2.E.- EL PROBLEMA DEL SESGO POR ERROR DE MEDIDA EN LOS DATOS DE PANEL

El sesgo que cometemos ignorando la correlación entre "η_i" y "x_{it}", al no aplicar el estimador IG viene definido precisamente por el tamaño de la covarianza que mide esa relación con respecto a la variación de la "x":

$$\beta_{SESGADO} = \beta + \frac{Cov(x_{it}, \eta_i)}{V(x_{it})}$$

Sin embargo, y por otro lado, el sesgo por error de medida se amplifica en el caso de estimar un modelo en diferencias (**Intragrupos en el caso de T=2**) respecto al caso en niveles y lo mismo ocurre, aunque en menor medida, con un modelo en desviaciones respecto a las medias o en desviaciones ortogonales (**Estimador Intragrupos**). La razón proviene de la forma del sesgo por error de medida:

EN NIVELES	EN DIFERENCIAS
$\dfrac{\beta}{1+\frac{V(\varepsilon)}{V(x^*)}}$	$\dfrac{\beta}{1+\frac{V(\Delta\varepsilon)}{V(\Delta x^*)}}=\dfrac{\beta}{1+\frac{(V\varepsilon_t+V\varepsilon_{t-1}-2Cov\varepsilon_t\varepsilon_{t-1})}{(Vx_t+Vx_{t-1}-2Covx_t x_{t-1})}}=\dfrac{\beta}{1+\frac{(2V\varepsilon-2Cov\varepsilon_t\varepsilon_{t-1})}{(2Vx-2Covx_t x_{t-1})}}$
	A poco que la Cov($x_t x_{t-1}$) sea mayor que la Cov($u_t u_{t-1}$), cosa muy lógica, el sesgo crece con relación al caso en niveles.

donde:

$$x = x^* + \varepsilon$$

Al usar desviaciones ortogonales o a la media, esas covarianzas se desvanecen ya que no se tiene en cuenta la Cov ($x_t x_{t-1}$) sino Cov (x_t,Media....) por lo que el sesgo tiende al mismo valor que el caso en niveles.

Si combinamos los dos puntos anteriores puede plantearse un problema: si al pasar de una estimación en niveles a otra con IG el resultado es muy diferente ¿cómo sabemos si se trata de una reducción del sesgo de efectos fijos por usar IG o una amplificación de un sesgo de error de medida ya presente en la ecuación en niveles que se ha amplificado mucho al usar IG?. *La solución es sencilla y pasa por comparar IG con 1^{as} diferencias: la estimación por MCO en primeras diferencias no está sesgada por efecto de la correlación entre η_i y x_{it}, otra cosa es que tenga autocorrelación y por tanto no sea eficiente, pero el valor del parámetro debe ser similar al IG. Si no es así, una posible razón puede ser la existencia de un sesgo por error de medida ya que la "amplificación si es mucho mayor en 1^{as} diferencias que en IG*

En realidad, **la combinación de sesgos no siempre es "a la baja"**, y esto es peor aún, ya que en la medida en que la correlación entre "η" y "x" fuese positiva y β también, los sesgos tenderían a compensarse. Así, :

- **en una estimación en niveles**, los sesgos podrían compensarse y esto es bueno, porque reduciria el error de aplicar esta estimación cuando no se debe

- **en una estimación en diferencias** o, en menor medida IG, la corrección del sesgo por efectos fijos podría compensarse por el aumento de un sesgo previo por error de medida sin que se notaran diferencias entre una estimación MCO en niveles y una IG.

Error de medida / **Efectos Fijos**	**SI**	**NO**
SI	- La estimación **en niveles** podría compensar los sesgos si Cov(η,x)>0 y β>0. Pero si estos dos signos son diferentes el sesgo seguiría existiendo al alza (para β<0) y a la baja para (β>0). - La estimación en **diferencias** o **IG** solucionará todo sesgo de efectos fijos pero muy probablemente exacerbaría el sesgo por error de medida.	- La estimación **en niveles** estará sesgada al alza *(si Co(η,x)>0)* o a la baja *(si Co(η,x)<0)*. - La estimación en **diferencias** o **IG** solucionará todo sesgo.
NO	- La estimación en **niveles** presentaría el sesgo por error de medida. - La estimación en **diferencias** o **IG** sólo serviría para exacerbar el sesgo por error de medida.	- La estimación en **niveles** no presentaría ningún sesgo. - La estimación en **diferencias** o **IG** no presentaría ningún sesgo si bien tampoco mejoraría la estimación en niveles.

BIBLIOGRAFÍA BÁSICA

Arellano, M. y Bond, S. (1988ª). "Some Test of Especification for Panel data: Monte Carlo Evidence and and an Application to Employment Equiations". *Applied Economics Discussion Papers 55*, Oxford.

Arellano,M. y Bover, O. (1990). "La Econometría de Datos de Panel". *Investigaciones Económicas*. Vol. 14. Nº 1. pp. 3-45.

Deaton, A. (1985). "Panel data from Times Series of Cross Sections". *Journal of Econometrics*, 30. pp. 109-126.

Chamberlain, G. (1984). "panel data". Z. Griliches and M.D. Intrilligator (editores). Handbook of Econometrics, Vol. II. Elsevier Science.

Griliches, Z. (1977). "Estimating the returns of Schooling: some econometric problems". *Econométrica*, 45, pp. 1- 22.

Hausman, J. A. (1978). "Specification tests in Econometrics". *Econometrica*, 46, pp. 1251-1272.

Johnston, J. (1992). "Métodos de Econometría". Ed. Vicens Vivens.

MaCurdy, T.E. (1981). "An Empirical Model of Labor Supply in a Life-Cycle Setting". *Journal of Political Economy*, 89, pp.1059-1085.

Mundlak, Y. (1978). "On the pooling of time series and cross section

Modelos de variables discretas y variables dependientes limitadas: modelos logit y probit.

- **Modelos de elección simple y múltiple. Regresión logit y probit. Modelos multilogit y multiprobit.**

J. Muro
Universidad de Alcalá
febrero 2008

Sílabo.

Modelos de elección discreta.

1. Modelos dicotómicos.

    * *Modelo lineal de probabilidad*.
    * *Modelo probit*.
    * *Modelo logit*.

2. Modelos multinomiales.

    2.1. Modelos para datos no ordenados.

    * *Modelo logit (probit) multinomial*.
    * *Modelo logit (probit) condicional*.

    2.2. Modelos para datos ordenados.

    * *Modelo logit (probit) ordenado*.
    * *Modelo de variable dependiente agrupada*.

    2.3. Modelos anidados.

.

Modelos de elección discreta.

Se analiza brevemente la construcción de modelos en los que la variable del lado izquierdo toma valores discretos, es decir, no puede alcanzar cualquiera de los valores contenidos en un intervalo concreto. Los valores observados pueden representar tanto la identificación de decisiones alternativas, que los agentes económicos pueden tomar, como el número de veces que acontece un suceso. Los modelos dicotómicos son los que presentan mayor simplificación entre los modelos de elección discreta. No obstante, ya que una mera alternativa dicotómica presenta los rasgos característicos de este tipo de modelos, los modelos con dos alternativas son utilizados ampliamente con fines didácticos.

1. Modelos dicotómicos.

Los modelos dicotómicos se caracterizan porque la variable del lado izquierdo no sólo es discreta sino dicotómica. Sólo puede tomar dos valores que, en general, se asocian a los valores uno y cero. El primero de ellos se utiliza habitualmente para representar la toma de una decisión y, el segundo, para indicar que dicha decisión no se ha adoptado.

Ejemplos de análisis dicotómicos podemos encontrar en el estudio de cuestiones tales como el consumo de bienes duraderos; la toma de decisiones empresariales como la contratación y despido de trabajadores; las intenciones o resultados del voto; y en temas muy frecuentes de economía laboral: actividad o inactividad, ocupación o paro, irregularidad o regularidad en el empleo, etc.

Planteamiento del modelo.

Los datos observados nos dicen que disponemos de una variable, **y,** a estudio, representativa de la toma de decisiones o de la ocurrencia de un suceso, que sólo toma los valores 1 y 0. También tenemos información de un conjunto de variables o características **X** que influyen sobre la probabilidad de que la variable **y** tome dichos valores. El modelo empírico de la variable **y** depende, por tanto, de los valores de las variables incluidas en **X,** y de un conjunto de parámetros que denominaremos $\boldsymbol{\beta}$.

En la literatura se enfoca la construcción de estos modelos desde dos puntos de vista:

a. Construcción del modelo a partir de una forma reducida.

b. Modelización de la probabilidad asociada a la variable dicotómica a través de las expresiones derivadas de una formulación estructural.

A continuación, desarrollamos en detalle ambos enfoques para situar en su contexto los bien conocidos modelos: *lineal de probabilidad, logit y probit*.

a. Construcción del modelo a partir de una forma reducida.

Para la construcción de un modelo dicotómico a partir de una forma reducida lo que habitualmente se hace es seguir el espíritu de un modelo de regresión, con la única diferencia de que la variable del lado izquierdo tiene sus valores restringidos y sólo toma los valores uno y cero. En la ecuación que se especifique en cada caso concreto habrá las correspondientes variables del lado derecho y un conjunto de parámetros.

Se parte del supuesto de que la probabilidad que buscamos es una función de la matriz de características **X** y de los parámetros a estimar. Bajo la hipótesis de linealidad, ampliamente utilizada en econometría, se considera que el argumento de la función anterior es una combinación lineal de las características y de los parámetros. Así, la probabilidad de que ocurra un suceso, $\mathbf{P_i}$, y de que éste no ocurra pueden expresarse como

$$P_i = Prob[\, y_i = 1 \mid X, \beta\,] = G(\, X_{i'}, \beta\,) = G(\, X_{i'} \beta\,). \qquad [1.1]$$

$$Prob[\, y_i = 0 \mid X, \beta\,] = 1 - G(.). \qquad [1.2]$$

A partir de este momento, por motivos de simplicidad en la notación, en general, se omitirá la expresión de condicionamiento en las fórmulas de probabilidades y momentos de las variables, que se dan por incluidos. Aunque estemos principalmente interesados en la expresión de las probabilidades, a partir de los resultados en (1.1) y (1.2) podemos conocer la esperanza de la variable a estudio, $\mathbf{Y_i}$, ya que

$$E(\, y_i\,) = 1 * P_i + 0 * (1 - P_i\,) = P_i\,. \qquad [1.3]$$

Como puede verse en (1.3), la probabilidad de que ocurra el suceso coincide con la esperanza de la variable dicotómica.

Parece lógico, en principio, suponer que las variables incluidas en **X** son no aleatorias, constantes en muestras repetidas o, en el caso extremo, sus DGP tienen parámetros que no influyen en la estimación de los parámetros relevantes en nuestro modelo. En otras palabras, consideramos al menos que **X** es débilmente exógena.

¿Cuáles pueden ser las candidatas para la función G(.)?

Por seguir con el espíritu del modelo de regresión, la primera candidata es la función identidad, es decir G(.) = 1. En este caso, la construcción del modelo da origen al *modelo lineal de probabilidad.* Su expresión formal es

$$y_i = X'_i\, \beta + \varepsilon_i\,. \qquad \forall i = 1,2,......N \qquad [1.4]$$

donde suponemos que la distribución de la perturbación aleatoria tiene esperanza nula como es habitual en los modelos de regresión.

Otra forma alternativa de ver lo anterior es a través de la siguiente transformación

$$y_i = E(\, y_i\,) + [\, y_i - E(\, y_i\,)]$$
$$y_i = G(X'_i\, \beta\,) + \varepsilon_i$$
$$y_i = X'_i\, \beta + \varepsilon_i\,.$$

¿Qué otras candidatas se conocen en la literatura? Para poder sugerir otras candidatas para la función G(.), debemos fijarnos en ciertos elementos del problema a estudio. En primer lugar, los valores de la probabilidad deben quedarse restringidos al intervalo 0-1. En segundo lugar, la probabilidad debe ser una función monótona de los argumentos de la función, es decir, parece plausible que cuando los argumentos tiendan a infinito, la probabilidad tienda a 1 y que cuando los argumentos tiendan a menos infinito, la probabilidad tienda a cero. Si tenemos en cuenta estas dos condiciones parece evidente que G(.) debe ser una función de distribución. Sin que otras candidatas se hayan vetado expresamente, las utilizadas en la literatura han sido la función de distribución Normal y la Logística. La primera de ellas da lugar al modelo *probit* y la segunda al modelo *logit.*

La razón para que ambos modelos hayan ido de la mano tanto tiempo se encuentra en que ambas distribuciones son muy parecidas excepto en el tamaño de las colas. Las colas de la función de distribución Logística son de mayor volumen. Cabe decir que la distribución Logística se parece bastante a una t de Student con 7 grados de libertad **.**

La expresión formal del modelo probit es

$$P_i = Prob(y_i = 1) = \int_{-\infty}^{X_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = F(X'_i \beta). \quad [1.5]$$

La formulación del modelo logit es

$$P_i = Prob(y_i = 1) = \frac{1}{1 + e^{-X'_i\beta}} = \Lambda(X'_i \beta). \quad [1.6]$$

Como puede verse en las fórmulas en (1.5) y (1.6) reservamos F(.) para la función de distribución normal, N(0,1), como es habitual, y Λ(.) para la función de distribución logística.

Cabría señalar que en estos dos últimos modelos los efectos marginales de las características sobre la variable a estudio no coinciden (como en las regresiones y, por tanto, en el modelo lineal de probabilidad) con los parámetros del modelo. Sobre este tema se profundizará más adelante.

b. Modelización del DGP a través de las expresiones derivadas de una formulación estructural.

En la literatura se adoptan dos enfoques para la forma estructural subyacente a un modelo dicotómico:

a. Modelización a través de una función índice (la forma estructural es el modelo de una variable latente).

b. Modelización a través de la formulación de una utilidad aleatoria.

Bajo el enfoque **a**, la variable observada toma unos valores que responden al comportamiento de una variable índice (latente o inobservable). Si el índice supera un determinado nivel la variable discreta toma el valor uno y, si no lo supera, toma el valor cero.

La variable latente está relacionada con las características a través de un modelo de regresión.

$$I_i^* = X'_i \beta + \varepsilon_i, \quad [1.7]$$

donde I^*_i representa a la variable latente, inobservable, y los demás elementos son los habituales en los modelos de regresión.

Según establezcamos el supuesto de que la perturbación aleatoria se distribuya como una Normal o una Logística, el modelo generado a partir de la expresión en (1.7) será un probit o un logit, respectivamente.

Los valores de la variable a estudio, observados, se relacionan con los de la variable latente en la forma

$$y_i = 1 \;\; si \;\; I_i^* > 0;$$
$$y_i = 0 \;\; si \;\; I_i^* \leq 0.$$

En este enfoque cabe destacar ciertos elementos que señalamos a continuación.

El hacer la varianza del modelo igual a uno es el resultado de un mero proceso de normalización, ya que el modelo no se altera porque se multiplique por cualquier cantidad.

El umbral considerado a superar por el índice puede ser cero o cualquier otro valor. En ciertos estudios se sugiere que sea el término constante.

La probabilidad asociada a la realización del suceso tiene la expresión ya conocida porque

$$\begin{gathered} P_i = Prob(\, y_i = 1) = Prob(\, I_i^* > 0) = \\ Prob(X'_i\, \beta + \varepsilon_i > 0) = Prob(\, \varepsilon_i > -X'_i\, \beta\,) = \\ Prob(\, \varepsilon_i > -X'_i\, \beta_i) = G(X'_i\, \beta\,). \end{gathered} \qquad \textbf{[1.8]}$$

Bajo el enfoque **b** la decisión que adopta cada individuo depende de que la utilidad esperada del nuevo estado, la nueva situación elegida, supere la utilidad que le reporta el estado actual más los costes asociados a la transición entre estados.

En el supuesto de que la utilidad de un individuo sea estocástica será tratado en el apartado de modelos multinomiales.

El modelo lineal de probabilidad.

Como antes se ha dicho, en el modelo lineal de probabilidad la función G(.) se escoge como la función identidad. La expresión del modelo se encuentra en (1.4).

La forma del modelo lineal de probabilidad sugiere una estimación por mínimos cuadrados ordinarios (MCO) para obtener estimadores con buenas propiedades estadísticas. Sin embargo, la aplicación de MCO en (1.4) plantea una serie de dificultades entre las que están las siguientes:

1. Si la esperanza de las perturbaciones aleatorias es cero, dado que la variable del lado izquierdo sólo toma los valores cero y uno, la perturbación aleatoria no puede tener una distribución normal. Sólo puede tomar dos valores, con una probabilidad asociada determinada. Esto origina que las probabilidades puedan estar fuera del intervalo 0-1, e incluso que puedan ser negativas. La estimación restringida del modelo para obligar que las probabilidades estén entre 0 y 1, no soluciona el problema a la hora de predecir. En la predicción se pueden obtener valores superiores a 1, lo que va en contra de la definición de probabilidad.

En efecto,

$$\begin{aligned} &Si\ \ E(\varepsilon_i) = 0 \\ &\varepsilon_i = 1 - X'_i\,\beta \ \ para\ \ y_i = 1 \\ &\varepsilon_i = -X'_i\,\beta \ \ para\ \ y_i = 0. \end{aligned}$$

Con probabilidades asociadas iguales a

$$\begin{aligned} &Prob(\varepsilon_i = 1 - X'_i\,\beta) = X'_i\,\beta; \\ &Prob(\varepsilon_i = -X'_i\,\beta) = 1 - X'_i\,\beta. \end{aligned}$$

2. Como consecuencia de lo anterior, los errores estándar y los estadísticos t estarán sesgados.

3. Además, las perturbaciones del modelo son heteroscedásticas. Por lo tanto, la matriz de varianzas y covarianzas estará mal calculada. Si calculamos la varianza de las perturbaciones aleatorias

$$\begin{aligned} Var(\varepsilon_i) &= (1 - X'_i\beta)^2 X'_i\beta + (-X'_i\beta)^2(1 - X'_i\beta) = \\ &(1 - X'_i\beta)[(-X'_i\beta)^2 + X'_i\beta - (X'_i\beta)^2] = \\ &P_i(1 - P_i) = E(y_i)(1 - E(y_i)). \end{aligned} \qquad [1.10]$$

Aplicaciones que todavía utilizan este modelo (cuyas estimaciones parecen estar relacionadas proporcionalmente con los modelos logit y probit, al menos en un intervalo de probabilidad relevante, como veremos más adelante) utilizan en la estimación de los parámetros del modelo el método de mínimos cuadrados generalizados factibles (MCGF). En este caso, las ponderaciones que se deben utilizar serían

$$w_i = \sqrt{\hat{y}_i(1 - \hat{y}_i)}. \qquad [1.11]$$

El proceso de estimación se llevaría a cabo en dos etapas, la primera para estimar ω_i y la segunda para estimar los parámetros de nuestro modelo por MCGF. Como se sabe, el procedimiento cabe iterarlo para obtener una mayor eficiencia para muestras pequeñas.

Modelos probit y logit.

La expresión de ambos modelos se encuentra respectivamente en (1.5), para el modelo probit, y en (1.6), para el logit. La expresión general de la función de verosimilitud para ambos modelos es

$$L(\beta \mid y_i, X'_i) = \prod_{i=1}^{N} G(X'_i\beta)^{y_i}[1 - G(X'_i\beta)]^{1-y_i}.$$

$$\ln L(\beta) = \sum_{i=1}^{N} \{y_i \ln G(.) + (1 - y_i)\ln[1 - G(.)]\}. \qquad [1.12]$$

Como puede verse, los valores de la variable iguales a uno, la realización de un suceso, contribuyen a la función de verosimilitud con la probabilidad de que ocurra el suceso y los valores de la variable iguales a cero, la no realización de un suceso, contribuyen con la probabilidad de que el suceso no ocurra.

Si calculamos las ecuaciones de verosimilitud para ambos modelos, su expresión es

$$\frac{\partial \ln L(\beta)}{\partial \beta} = 0;$$

$$\sum_{i=1}^{N}\{y_i \frac{g(.)}{G(.)}.X_i + (1 - y_i)\frac{g(.)}{1 - G(.)}.X_i\} = 0. \qquad [1.13]$$

$$\sum_{i=1}^{N}\frac{y_i - G(.)}{G(.)[1 - G(.)]}.g(.).X_i = 0.$$

Este sistema de ecuaciones no es lineal y debe resolverse por alguno de los conocidos métodos de optimización no lineal, por ejemplo, el de Newton-Raphson o el de Berndt-Hall-Hall-Hausman (BHHH).

Pasemos a concretar las dos expresiones generales en (11.12) y (1.13). La expresión de las ecuaciones de verosimilitud para el modelo probit es

$$\sum_{y_i=0}\frac{-f(.)}{1 - F(.)}.X_i + \sum_{y_i=1}\frac{f(.)}{1 - F(.)}.X_i = 0, \qquad [1.14]$$

donde f(.) es la función de densidad de probabilidad de una variable aleatoria normal, N(0,1).

Las ecuaciones de verosimilitud para el modelo logit son

$$G(.) = \Lambda(X'_i\beta); \;\; g(.) = \Lambda(.)[1 - \Lambda(.)].$$

$$\sum_{i=1}^{N}[y_i - \Lambda(.)].X_i = 0. \qquad [1.15]$$

Donde, recordamos que

$$\Lambda(X'_i \beta) = \frac{1}{1 + e^{-X'_i \beta}},$$

es la función de distribución logística.

En el modelo logit, si el conjunto de variables explicativas contiene una constante, se cumple que el promedio de las probabilidades predichas iguala al porcentaje de unos en la muestra.

Una interpretación conveniente, que se utilizará profusamente en la realización de contrastes de especificación, es que en la expresión (1.15) la diferencia contenida entre corchetes puede ser considerada un residuo generalizado. Esta interpretación también abre una vía a la estimación por el método generalizado de momentos (MGM) que utilizaremos ampliamente en estas lecciones.

Analicemos el Hessiano en ambos modelos para poder sacar conclusiones sobre el proceso no lineal de maximización de la función de verosimilitud.

Para el modelo probit

$$H = -\sum_i \lambda_i (\lambda_i + X'_i \beta) X_i X'_i .$$

$$\text{donde}$$

$$\lambda_i = \lambda_{0i} = \frac{-f(.)}{1-F(.)} \quad si \quad y_i = 0;$$

$$\lambda_i = \lambda_{1i} = \frac{f(.)}{F(.)} \quad si \quad y_i = 1.$$

Para el modelo logit

$$H = -\sum_i \Lambda(.)[1 - \Lambda(.)] X_i X'_i .$$

Ambos hessianos son definidos negativos lo que garantiza la convergencia en el cálculo del máximo de la función de verosimilitud.

Cálculo de los errores estándar de las probabilidades predichas y de los efectos marginales estimados.

La cuestión más interesante a plantear en este caso es que tanto las probabilidades predichas como los efectos marginales son funciones no lineales de los parámetros estimados. En consecuencia, los errores estándar de esas estimaciones deberán aproximarse mediante algún método oportuno. El más utilizado dentro de la literatura empírica es el denominado *método delta*.

La probabilidad predicha en los modelos probit y logit es

$$\tilde{G}(.) = G(X'_i\ \tilde{\beta}).$$

Se puede calcular su matriz de varianzas y covarianzas a través de la aplicación del *método delta*. Su resultado son las expresiones siguientes

$$var\ asint.(\tilde{G}) = \left(\frac{\partial \tilde{G}}{\partial \tilde{\beta}}\right)' V \left(\frac{\partial \tilde{G}}{\partial \tilde{\beta}}\right).$$

$$donde$$

$$V = var\ asint.(\ \tilde{\beta}\ ).$$

$$Expresión\ que\ es$$

$$var\ asint.(\tilde{G}) = [\tilde{g}(.)]^2\ X'_i\ VX_i\ .$$

La expresión general de los efectos marginales en los modelos probit y logit es

$$\frac{\partial E(\ y_i\ )}{\partial\ X_i} = \frac{\partial G(X'_i\ \beta\ )}{\partial (X'_i\ \beta\ )}|_{\beta=\tilde{\beta}}\ .\tilde{\beta} = g(X'_i\ \beta\ )|_{\beta=\tilde{\beta}}\ .\tilde{\beta}.$$

Que se convierte en el modelo logit en

$$\frac{\partial E(\ y_i\ )}{\partial\ X_i} = \Lambda(X'_i\ \beta\ )(1 - \Lambda(X'_i\ \beta\ )\beta = P_i\ (1 - P_i\ )\beta.$$

También se puede calcular la matriz de varianzas y covarianzas de los efectos marginales, *método delta*. por medio de

$$Var\ asint.(g(X_i\ \beta\ )|_{\beta=\tilde{\beta}}\ .\tilde{\beta}\ ) = \left(\frac{\partial \tilde{\gamma}}{\partial \tilde{\beta}}\right)' V \left(\frac{\partial \tilde{\gamma}}{\partial \tilde{\beta}}\right).$$

$$donde$$

$$V = var\ asint.(\tilde{\beta}\ );$$

$$\tilde{\gamma} = g(X'_i\ \beta\ )|_{\beta=\tilde{\beta}}\ .\tilde{\beta}.$$

Conviene destacar que además del concepto de elasticidad, en los estudios aplicados con modelos dicotómicos se suele emplear el concepto de seudoelasticidad o simplemente *elasticidad*. Se llama así a la alteración relativa que se produce en la probabilidad de que ocurra un fenómeno cuando una variable dicotómica de las incluidas en el lado derecho de la ecuación pasa de tomar el valor 1 al 0 o viceversa. Un ejemplo típico es el cálculo de la *elasticidad* de la toma de una decisión frente al género. La expresión que permite el cálculo de esta *elasticidad* es muy simple y se plantea como objetivo para el lector.

Inferencia estadística: Contrastes.

La realización de contrastes en los modelos dicotómicos se lleva a cabo por los métodos bien conocidos en la literatura econométrica. Dentro de éstos, cada vez se populariza más el uso de contrastes de momentos condicionales (CMC). Aquí, reseñaremos algunos casos concretos de contrastes de especificación.

1. Variables omitidas.

Las hipótesis nula y alternativa en este caso son

$$H_0 : y = X_1'\beta_1 + \varepsilon.$$
$$H_1 : y = X_1'\beta_1 + X_2'\beta_2 + \varepsilon$$

El contraste a realizar es del tipo de multiplicadores de Lagrange. Su forma es TR^2, donde el R^2 se obtiene de la regresión auxiliar (el denominador del R^2 sin centrar) de r_i frente a X^*_i. Estas variables auxiliares se construyen por medio de las expresiones siguientes

$$r_i = y_i \left[\frac{1 - G(.)}{G(.)} \right]^{1/2} + (y_i - 1) \left[\frac{G(.)}{1 - G(.)} \right]^{1/2} .$$

$$X_i^* = \frac{g(.)}{\left[G(.)(1 - G(.)) \right]^{1/2}} \cdot X_i .$$

2. Heteroscedasticidad.

Para el análisis de la heteroscedasticidad se acostumbra a utilizar un contraste del tipo de heteroscedasticidad multiplicativa, Harvey(1976). Las hipótesis nula y alternativa son

$$H_0 : var(\varepsilon) = \sigma^2 I.$$
$$H_1 : var(\varepsilon) = \left[\exp(\gamma' z) \right]^2 .$$

Como sugiere Harvey(1976) se aplica un contraste de multiplicadores de Lagrange.

Davidson y McKinnon(1984) han demostrado que este último contraste no es apropiado para el modelo logit y si para el probit. En el ejemplo se desarrolla la forma de realizar este contraste de hetroscedasticidad.

Bondad del ajuste.

La bondad del ajuste en los modelos de elección discreta es un tema controvertido, ya que en estos modelos no existe una interpretación de las medidas de la bondad del ajuste tan intuitiva como en el modelo clásico de regresión. Cragg y Uhler(1970), Amemiya(1981), Maddala(1983), McFadden(1974) y Zavoina y McElvey(1975), entre otros, son ejemplos de la preocupación por el estudio de esta cuestión.

Como mínimo, se puede establecer que en la presentación de las estimaciones se deben presentar los valores del logaritmo de la función de verosimilitud del modelo estimado y del ingenuo (modelo con sólo una constante).

Seudo R^2. Se denomina así a una medida de la bondad del ajuste que tiene como expresión

$$1 - \frac{\ln L}{\ln L_0}.$$

$$\ln L_0 = N[P \ln P + (1 - P)\ln(1 - P).$$

Tabla de predicciones. Se suele presentar una tabla 2x2 que recoge el cruce entre los valores observados de la variable a estudio y de las predicciones del comportamiento de dicha variable obtenidas con nuestro modelo bajo la regla siguiente

$$\hat{P}_i > 0.5 \text{ implica que } \hat{y}_i = 1;$$

$$P_i \leq 0.5 \text{ implica que } \hat{y}_i = 0.$$

Hay que tener cuidado en la utilización de las tablas de predicciones cuando la muestra disponible está desequilibrada, ya que el criterio del 0.5 lleva a unas predicciones con un alto grado de inexactitud.

Sugerencias respecto a econometría aplicada.

La primera de las sugerencias que cabe apuntar se refiere a las predicciones que se obtienen de los modelos probit y logit. En general son muy parecidas, pero difieren en los casos siguientes:

Si el modelo es desequilibrado: es decir, hay pocos unos o pocos ceros.

Si la variabilidad de una de las características relevantes al problema que estamos estudiando es muy grande.

La segunda concierne al cálculo de los efectos marginales. Como ya hemos visto, los efectos marginales dependen de los valores de la matriz X. Deben calcularse para las medias de los regresores y otros parámetros de interés.

La tercera y última se refiere a las varianzas de los modelos probit y logit y sus consecuencias para la comparación entre los coeficientes de los modelos analizados. En el modelo probit la varianza de la distribución es 1, ya que se normaliza. En el modelo logit, esta varianza es $\pi^2/3$. Así, para comparar la magnitud de los coeficientes de ambos modelos deberíamos multiplicar los coeficientes del probit por 1.8. Ya veremos más adelante que es lo que sugiere Amemiya(1981).

2. *Modelos multinomiales*.

Se suelen denominar modelos multinomiales a aquellos en que la variable a explicar es discreta, con una gama de valores de tamaño superior a dos. Esta denominación abarca un conjunto de modelos dispares ya que en numerosas ocasiones se obtienen datos que exigen la formulación de modelos multinomiales. Ejemplos de la abundancia de estos modelos pueden ser el estudio de las alternativas en el uso del transporte, McFadden(1974), Hensher(1986); de la elección de ocupación por parte de las personas ocupadas, Boskin(1974), Schmidt y Strauss(1975a y 1975b); la asignación de una valoración a obligaciones empresariales, Terza(1985); los análisis de las encuestas de opinión; los resultados del voto; los grupos de renta; la posición de un individuo en el mercado de trabajo, etc.

En la información que proporcionan los datos de los estudios citados en el párrafo anterior hay un elemento importante a considerar que es la naturaleza de los propios datos. El resultado de la observación de una variable discreta con múltiples alternativas puede incluir bien la alternativa elegida bien el dato de ésta última y una ordenación de las preferencias del individuo.

Cuando los datos a estudio facilitan la alternativa elegida y una ordenación de las distintas alternativas posibles diremos que los datos son ordenados. Como ejemplos podemos poner las encuestas de opinión, los votos, los grupos de renta, etc. En el caso de que sólo dispongamos de la información sobre la alternativa elegida diremos que los datos son no ordenados. Como ejemplos pueden citarse el uso del transporte, elegir una ocupación en el mercado de trabajo, etc.

Los modelos que utilizaremos para el análisis de los dos tipos de datos, ordenados y no ordenados, serán distintos ya que la introducción de toda la información que contienen los datos en el proceso de modelización aumenta la eficiencia de las estimaciones. Para los datos ordenados los modelos serán más parecidos a los de regresión y tenderán a éstos cuando el número de las alternativas sea muy grande. Para los datos no ordenados, que por otra parte son muy frecuentes, los modelos empleados son generalizaciones de los probit y logit dicotómicos.

2.1. Modelos para datos de elección entre múltiples alternativas no ordenadas.

Como se ha dicho antes, este tipo de modelos son generalizaciones de los modelos dicotómicos ya estudiados. En función del tipo de variables que podamos utilizar para determinar la probabilidad de elección de una alternativa clasificaremos los modelos en multinomiales y condicionales.

Planteamiento del modelo.

Para abordar el planteamiento de estos modelos introduzcamos una notación conveniente. Realicemos, en primer lugar, una revisión de la información disponible. La muestra proviene de N agentes económicos que toman una decisión elegida entre un conjunto de **J** alternativas.

Cada uno de los agentes económicos tiene un conjunto de características, variables que califican a cada individuo, que forman para cada agente un vector de dimensión *K*, al que llamaremos w_i. Estas características determinan la elección de una alternativa y son constantes para cualquier alternativa que elija un agente económico. Es decir,

$$w_i \text{ es } (Kx1) \text{ para cada individuo;}$$
$$W = \{ w_i' \} \text{ es } (NxK) \text{ y reune las características de todos los individuos.}$$

Las alternativas también gozan de un conjunto de atributos, que son bien variables objetivas de las alternativas, bien aspectos subjetivos de las alternativas para cada agente económico, que forman para cada alternativa y agente económico un vector de dimensión **L**, al que llamamos X_{ij}. Estos atributos influyen también en la probabilidad de elegir una alternativa. Debemos insistir que los atributos pueden ser comunes para todos los agentes en la muestra, pero también pueden ser específicos para cada agente y alternativa. Es decir,

X_{ij} es (Lx1) para cada alternativa;
X = { X_{ij}´} es (NJxL) y reúne los atributos de las J alternativas.

En el caso general, tanto las características como los atributos son el conjunto de las variables del lado derecho de la ecuación y forman un vector que llamamos Z_{ij} de dimensión **L+K**. Es decir,

Z_{ij} es ((L+K)x1) para cada individuo i y alternativa j.

Una vez completadas las cuestiones de notación, podemos abordar el planteamiento del modelo. La construcción del modelo empírico se lleva a cabo a través de la formulación de una utilidad aleatoria. Para cada agente económico i, la alternativa j le proporciona una utilidad U_{ij}, aleatoria, que puede formularse como

$$U_{ij} = S_{ij} + \varepsilon_{ij}, \qquad \textbf{[1.16]}$$

donde,

$$S_{ij} = Z_{ij}\gamma; \quad \gamma = [\beta'\alpha'].$$

El agente económico elegirá la alternativa que le reporte la utilidad máxima. Por lo tanto:

$$\text{Prob } (y_i = j) = \text{Prob } (U_{ij} > U_{ik}) \quad k = j.$$

En la anterior especificación en (1.16), el supuesto que se establezca sobre la distribución del término de error determinará que el modelo sea un probit (distribución normal) o un logit (distribución logística).

Conviene señalar desde ahora las dificultades que pueden plantearse en la estimación de modelos **probit**. La formulación teórica de un probit multinomial no plantea problemas. Su tratamiento entraña evaluación de integrales múltiples de la distribución normal que hasta el momento son difíciles de calcular (a no ser por métodos de simulación) y muy costosas en tiempos de ordenador. Por esta razón no se suele utilizar. Cuando se ha utilizado con un número de alternativas reducido, por ejemplo tres, no hay diferencias con los resultados obtenidos por los modelos logit.

Expresión formal de un modelo ***logit*** *para datos no ordenados.*

Si la función de distribución de las perturbaciones aleatorias en (1.15) es

$$G(\varepsilon_{ij}) = \exp(-e^{-\varepsilon_{ij}}). \ \forall i, j \qquad \textbf{[1.17]}$$

Es decir, la función de distribución de Weibull, y las perturbaciones se distribuyen independientemente, McFadden(1973) ha demostrado que la probabilidad de que el agente económico i escoja la alternativa j viene dada por

$$Prob(\, y_i \,=\, j) = \frac{e^{Z'_{ij}\gamma}}{\sum_j e^{Z'_{ij}\gamma}}. \qquad [1.18]$$

Antes hemos distinguido entre características y atributos. Debemos apreciar que si en la expresión anterior las separamos queda

$$Prob(\, y_i \,=\, j) = \frac{e^{X'_{ij}\beta + w_i\alpha}}{\sum_j e^{X'_{ij}\beta + w_i\alpha}} = \frac{e^{X'_{ij}\beta}}{\sum_j e^{X'_{ij}\beta}}. \qquad [1.19]$$

Es decir, al separar características y atributos desaparecen en la expresión del modelo las características de los individuos permaneciendo exclusivamente los atributos de las alternativas. Si queremos que las primeras no se pierdan tendremos que construir una variable ficticia que represente las alternativas y multiplicar por ella las características individuales. Con esto se resuelve el problema de la determinación de la influencia de los aspectos individuales, pero con la contrapartida de aumentar notablemente el número de variables en el lado derecho del modelo, hecho que puede afectar la estimación del modelo si el tamaño de la muestra no es excesivamente grande. Pongamos un ejemplo de esta reformulación del modelo.

Supongamos que las alternativas de un individuo son estar en la inactividad, ocupado o parado. Las características individuales de los individuos de la muestra son la edad y el sexo. Al introducir las variables ficticias de cada alternativa multiplicando a las características individuales pasamos de tener dos variables a seis variables que en términos cualitativos serían

v1	v2	v3	v4	v5	v6
edad	0	0	sexo	0	0
0	edad	0	0	sexo	0
0	0	edad	0	0	sexo

Por último, podemos apreciar que si la influencia de todos los atributos es despreciable la probabilidad de que un agente tome una decisión concreta es igual al recíproco del número de alternativas.

Modelo logit multinomial.

Se denomina así al modelo en el que las variables del lado derecho son exclusivamente las características de los agentes económicos. Este modelo es el que se suele utilizar con más frecuencia en los trabajos aplicados.

Dado que las variables son constantes para cualquier alternativa no podríamos apreciar los efectos individuales de cada alternativa si no introdujéramos una variable ficticia que represente cada una de dichas alternativas.

La expresión formal del modelo es

$$Prob(\ y_i\ =\ j) = P_{ij} = \frac{e^{w'_i\alpha_j}}{\sum_k e^{w'_i\alpha_k}}, \qquad [1.20]$$

donde k representa el índice que corre para las alternativas y el vector de parámetros lleva el subíndice correspondiente a la alternativa concreta analizada.

En el modelo logit multinomial existe una indeterminación de los parámetros, ya que si hacemos la transformación siguiente

$$\alpha_j^* = \alpha_j + \alpha_0.$$

y sustituimos en la expresión (1.19), las probabilidades facilitadas por el modelo no se alteran.

Para resolver esta indeterminación se introduce una normalización consistente en hacer α_0=0. Con esta convención la expresión en (1.19) del modelo logit multinomial queda

$$Prob(\ y_i\ =\ j) = \frac{e^{w'_i\alpha_j}}{1 + \sum_{k\neq 0} e^{w'_i\alpha_k}}. \qquad [1.21]$$

El modelo logit multinomial tiene una característica importante que ha hecho correr ríos de tinta porque afecta notablemente a la flexibilidad del modelo. Si calculamos la razón entre las probabilidades de aceptar una pareja de alternativas resulta la expresión

$$\frac{P_{ij}}{P_{i0}} = e^{w'_i\alpha_j}; \quad \frac{P_{ij}}{P_{ik}} = \frac{e^{w'_i\alpha_j}}{e^{w'_i\alpha_k}}.$$

Si aplicamos logaritmos queda

$$\log\left(\frac{P_{ij}}{P_{i0}}\right) = w'_i\alpha_j; \quad \log\left(\frac{P_{ij}}{P_{ik}}\right) = w'_i(\alpha_j - \alpha_k). \qquad [1.22]$$

Como puede verse la relación entre las probabilidades de decidir entre dos alternativas no depende del resto de las alternativas. En la expresión en (1.22) los únicos parámetros que afectan a la relación son los correspondientes a las alternativas relacionadas. Esta propiedad se conoce como la **independencia de las alternativas irrelevantes**. También es conocida como el problema de elección entre coche rojo o coche azul. Es una propiedad muy interesante en la estimación del modelo, pero provoca grandes problemas en la interpretación de los resultados.

Estimación del logit multinomial.

Si hacemos d_{ij}= 1 si y_i= j y d_{ij}= 0 si y_i no es igual a j, podemos abordar la construcción de la función de verosimilitud del modelo logit multinomial. Su expresión es

$$L(\alpha_j, j = 1,2....J) = \prod_{i=1}^{N}\prod_{j=1}^{J}[Prob(d_{ij} = 1)]^{d_{ij}};$$

$$\ln L = \sum_i \sum_j d_{ij} \ln\left(\frac{e^{w'_i \alpha_j}}{1+\sum_k e^{w'_i \alpha_k}}\right). \qquad [1.23]$$

Bondad del ajuste.

Como elemento básico se utilizan los valores de la función de verosimilitud del modelo contemplado y del modelo ingenuo.

El seudo R^2 también puede emplearse como una medida de la bondad del ajuste. Su expresión es

$$1 - \frac{\ln L}{\ln L_0}.$$

$$\ln L_0 = \sum_j N_j \ln P_j,$$

donde P_j es la proporción de sucesos j en la muestra.

La tabla de predicciones es otro índice de la bondad del ajuste del modelo logit multinomial. El criterio de construcción de la tabla es asignar el valor de la alternativa j siempre que la probabilidad predicha asociada a dicha alternativa sea superior a las asignadas a cualquiera otra alternativa.

Contrastes de especificación.

Hemos visto que un condicionante del modelo logit multinomial es la independencia de las alternativas irrelevantes. Dado que el supuesto es bastante fuerte a la vista de los datos en general disponibles, el contraste de esta condición se convierte en punto fundamental a la hora de saber si la especificación del modelo es o no correcta.

Hausman-McFadden(1984) han desarrollado un contraste de la independencia de las alternativas irrelevantes. Su expresión es

$$\chi^2 = (\hat{\beta}_s - \hat{\beta}_f)'[\hat{V}_s - \hat{V}_f]^{-1}(\hat{\beta}_s - \hat{\beta}_f) \tilde{H}_0\, \chi^2_q .$$

$$\begin{array}{c} \textit{donde} \\ \hat{\beta}_s : \textit{modelo restringido;} \\ \hat{\beta}_f : \textit{modelo } \text{sin} \textit{ restricciones;} \\ \hat{V}_s, \hat{V}_f : \textit{matrices de varianzas de ambos} \end{array} \quad [1.25]$$

Modelo logit condicional.

El modelo logit condicional responde a la situación en que las variables que influyen sobre la probabilidad de aceptar una alternativa son los atributos de las alternativas, que, como ya se ha comentado, pueden variar para cada individuo o no.

La expresión formal del modelo es

$$Prob(y_i = j) = \frac{e^{X'_{ij}\beta}}{\sum_j e^{X'_{ij}\beta}}. \quad j = 1,2,....J. \quad [1.26]$$

Como se ve, en la expresión existe un sólo vector de parámetros a estimar.

En el modelo logit condicional también se da la independencia de las alternativas irrelevantes. En efecto,

$$\log\left(\frac{P_{ij}}{P_{ik}}\right) = (X'_{ij} - X'_{ik})\beta. \quad [1.27]$$

La función de verosimilitud de este modelo es

$$\ln L = \sum_i d_{ij} \ln Prob(d_{ij} = 1). \quad [1.28]$$

Las ecuaciones de verosimilitud tienen la forma

$$\sum_i \sum_j P_{ij}(X_{ij} - \overline{X}_i), \quad [1.29]$$

donde,

$$\overline{X}_i = \sum_j P_{ij} X_{ij} .$$

El hessiano tiene la expresión

$$\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\sum_i \sum_j P_{ij}\left(X_{ij} - \overline{X}_i \right)\left(X_{ij} - \overline{X}_i \right)'.$$

La matriz de varianzas y covarianzas asintótica de los estimadores cabe obtenerla del Hessiano.

Bondad del ajuste.

Merece destacarse que si calculamos un seudo-R^2, en el modelo logit condicional hay dos posibles modelos restringidos. Uno de ellos es el resultante de hacer todos los parámetros nulos. Es decir

$$\beta = 0. \quad \ln L_0 = -N \ln J.$$

La segunda posibilidad es considerar un modelo ingenuo en el que se colocaran sólo un conjunto de variables ficticias que representen las J alternativas. Es decir,

$$\ln L_0 = \sum_j N_j \ln P_j.$$

Contrastes de especificación.

En este caso, el contraste fundamental es el de la independencia de las alternativas irrrelevantes, que como ya vimos antes es del tipo del sugerido por Hausman y McFadden(1984).

2.2. Modelos para datos de elección entre múltiples alternativas ordenadas.

Este tipo de modelos se construye en el caso en que los datos den información de dos cuestiones: de la alternativa escogida y de la ordenación de las alternativas posibles. Lo relevante en el análisis de la información es conocer si el límite del intervalo que determina la observación discreta es conocido (variable dependiente agrupada) o desconocido (por lo tanto a estimar en el modelo) por ser cualitativo.

Planteamiento del modelo.

Abordamos la modelización de la información discreta y ordenada a través de una función índice inobservable. La expresión es

$$y_i^* = X'_i \beta + \varepsilon_1. \qquad \textbf{[1.30]}$$

La conversión de los valores de la función índice en los valores discretos ordenados sigue la regla siguiente

$$
\begin{aligned}
y &= 0 \ \ si \ \ y^* \leq 0, \\
y &= 1 \ \ si \ \ 0 < y^* \leq \mu_1, \\
y &= 2 \ \ si \ \ \mu_1 < y^* \leq \mu_2, \\
y &= J \ \ si \ \ \mu_{J-1} < y^*.
\end{aligned}
\qquad [1.31]
$$

Como puede verse, la conversión puede considerarse una forma de censura de los datos. Los datos en vez de facilitar información detallada censuran la información y sólo se conoce el intervalo en el cual está la observación concreta.

La expresión formal de un modelo **probit** (**logit**) para datos ordenados puede escribirse como los elementos de la función de densidad de probabilidad de los datos. Estos son,

$$Prob(y = 0) = Prob(y^* \leq 0) = G(-X\beta),$$

$$Prob(y = 1) = Prob(0 < y^* \leq \mu_1) = G(\mu_1 - X\beta) - G(-X\beta),$$

$$......$$

En el caso de que la función de distribución G(.) sea la normal o la logística, se obtiene el modelo probit y logit, respectivamente. Si los parámetros μ_j son conocidos el modelo apropiado es el de variables dependientes agrupadas y si dichos parámetros son desconocidos los modelos probit y logit ordenados son los oportunos.

Modelo probit (logit) ordenado.

Dado que los límites del intervalo son desconocidos podemos hacer estándar el modelo y suponer que las perturbaciones aleatorias del modelo son N(0,1).

Expresión formal del modelo.

Para la estimación empleamos el método de la máxima verosimilitud. La función de verosimilitud es

$$L(\beta, \mu_j) = \prod_{i=1}^{N}\prod_{j=1}^{J}\left[G(\mu_j - X'_i\beta) - G(\mu_{j-1} - X'_i\beta)\right]^{d_{ij}}.$$

$$\ln L = \sum_i\sum_j d_{ij} \ln\left[G(\mu_j - X'_i\beta) - G(\mu_{j-1} - X'_i\beta)\right]. \qquad [1.33]$$

Donde,

$$d_{ij} = 1 \quad si \quad \mu_{j-1} < y_i^* \leq \mu_j ,$$
$$d_{ij} = 0 \quad en \; el \; resto \; de \; los \; casos.$$

Modelo de variable dependiente agrupada.

En el caso de que los datos nos proporcionen también información sobre los límites de los intervalos el modelo que se construye recibe el nombre de modelo de variable dependiente agrupada. Al conocer los límites del intervalo, el modelo no se puede hacer estándar y los parámetros a estimar serán β y σ^2.

Expresión formal del modelo.

La función de verosimilitud del modelo es

$$\ln L = \sum_i \sum_j d_{ij} \ln\left[G\left(\frac{\mu_j - X'_i \beta}{\sigma}\right) - G\left(\frac{\mu_{j-1} - X'_i \beta}{\sigma}\right)\right]. \quad [1.34]$$

Donde las equivalencias con la expresión anterior son inmediatas.

2.3 Modelos anidados.

Hasta ahora, hemos considerado modelos multinomiales, logit multinomial y logit condicional, que descansan sobre el supuesto de independencia de las alternativas irrelevantes. Como sabemos, esta hipótesis es heroica en numerosas circunstancias del análisis empírico. Si desechamos por el momento las alternativas multivariantes, que se acomodan bien a la relajación de este supuesto en el modelo logit multinomial, la especificación de modelos anidados se plantea como alternativa al incumplimiento del supuesto anterior en los modelos condicionales.

Un modelo anidado agrupa y jerarquiza las decisiones que puede adoptar un agente económico en una estructura de árbol. Hay decisiones que sólo pueden tomarse si previamente se han adoptado otras de una jerarquía mayor. En la construcción de modelos anidados, el diseño de un árbol de decisiones cuyos resultados sean los valores de la variable a explicar resulta indispensable. En ciertas ocasiones, este árbol vendrá dado por la propia naturaleza del problema; en otras, la estructura de nuestro árbol será establecida conforme al criterio específico del investigador. La teoría de grafos es un instrumento eficaz en esta tarea.

Desde un punto de vista estadístico, la comparación entre las especificaciones de un modelo condicional y un modelo anidado nos indica que el último considera las perturbaciones del modelo homoscedásticas, mientras que el primero considera heteroscedásticas las distintas ramas del árbol y sólo son homoscedásticas las arborescencias finales que parten de un mismo tronco. Pongamos un ejemplo. En la realización de un viaje se ofrecen las alternativas siguientes: avión, tren, autobús y coche. Si modelizamos estas alternativas con un modelo condicional, el cálculo de las probabilidades de utilizar un medio de locomoción concreto considera que las perturbaciones del modelo son homoscedásticas. Si, por el contrario, modelizamos las cuatro alternativas mediante un modelo anidado, en el que las primeras ramas sean medio público o privado, el cálculo de las probabilidades de elegir un medio de viaje considera que las perturbaciones de ambas ramas son heteroscedásticas, y deja la homoscedasticidad para los medios públicos.

Planteamiento del modelo.

Antes de abordar la especificación de un modelo anidado supongamos, como siempre, que disponemos de información sobre los atributos de las alternativas. En este caso, la información se encontrará jerarquizada y cabe dividirla en atributos de las ramas, z_l, y atributos de las alternativas incluidas en cada rama, X_{jl}. En un modelo anidado con dos niveles, *l* ramas con L_j alternativas en cada rama, la expresión de la probabilidad no condicionada de la alternativa *j* que pertenece a la rama *l* es

$$P_{jl} = \frac{e^{\beta' X_{jl} + \gamma' z_l}}{\sum_{l=1}^{L} \sum_{j=1}^{J_l} e^{\beta' X_{jl} + \gamma' z_l}}. \qquad [1.35]$$

Esta expresión puede escribirse como

$$P_{jl} = P_{j|l} P_l = \frac{e^{\beta' X_{jl}} e^{\gamma' z_l}}{\sum_{l=1}^{L} \sum_{j=1}^{J_l} e^{\beta' X_{jl} + \gamma' z_l}} \frac{\sum_j e^{\beta' X_{jl}} \sum_l e^{\gamma' z_l}}{\sum_j e^{\beta' X_{jl}} \sum_l e^{\gamma' z_l}}.$$

Definimos el valor inclusivo o de dentro de la rama como

$$I_l = \ln \sum_{j=1}^{J_l} e^{\beta' X_{jl}}. \qquad [1.36]$$

Como puede verse, el valor inclusivo recoge el sumatorio de todas las alternativas que cuelgan de la rama *l*.

A partir de las expresiones anteriores, y tras unas operaciones simples, podemos expresar la probabilidad de la alternativa *j* condicional a encontrarse en la rama *l* como

$$P_{j|l} = \frac{e^{\beta' X_{jl}}}{\sum_{j=1}^{J_l} e^{\beta' X_{jl}}}. \qquad [1.37]$$

y la probabilidad marginal de la rama *l* como

$$P_l = \frac{e^{\gamma' z_l + \pi_l I_l}}{\sum_{l=1}^{L} e^{\gamma' z_l + \pi_l I_l}}. \qquad [1.38]$$

Como puede verse, las probabilidades que suministra el modelo logit anidado, condicional y marginal, se comportan, respectivamente, como las de un modelo logit condicional interno para las alternativas de la rama correspondiente y como las de un logit condicional para el conjunto de las ramas. En este último, se consideran como atributos no sólo las z_l sino también los valores inclusivos I_l.

La generalización del modelo para varios niveles en el árbol incrementa enormemente su complejidad aunque las expresiones mantienen la misma estructura anterior. Ya que las estimaciones dependen de la estructura de árbol definida, la multiplicación de niveles aumenta también el número de modelos competitivos para el tratamiento del problema considerado.

Estimación.

La estimación de los parámetros del modelo anidado puede realizarse por el método de la máxima verosimilitud. Cabe emplear dos procedimientos que se relacionan a continuación.

1. *Estimación maximoverosímil con información limitada en dos etapas*.

El procedimiento se inicia con la estimación de modelos logit condicionales para cada rama. En segundo lugar, se estiman los modelos logit condicionales para todas las ramas con los atributos de rama y con las estimaciones de los valores inclusivos para cada rama obtenidos en (1.36).

Al ser el procedimiento bietápico debe corregirse la estimación de la matriz de varianzas y covarianzas en la segunda etapa.

2. *Estimación maximoverosímil con información completa*.

La función de verosimilitud del modelo logit anidado es

$$\ln L = \sum_{i=1}^{N} \ln [P_{j|l} \, P_l].$$

Cuya maximización permite obtener unos estimadores con mayor eficiencia que los obtenidos por el procedimiento anterior.

Debe señalarse que la estimación de los parámetros del modelo logit anidado depende de la estructura arbórea que proporcionemos a las alternativas. Una vez obtenidas las estimaciones de los modelos en competencia, por ahora, no hay métodos para contrastar cuál de estos modelos no anidados, en el sentido que ninguno de ellos es un caso particular del otro, es mejor.

BIBLIOGRAFIA.

Amemiya, T. (1981), "Qualitative Response Models: A Survey ". Journal of Economic Literature, 19, 4, págs. 481-536.

Amemiya, T. (1985), Advanced Econometrics. Harvard University Press.

Borjas, G.J. y G.T. Sueyoshi (1994) “A Two Stage Estimator for Probit Models with Structural Group Effects” Journal of Econometrics, 64, págs. 165-182.

Boskin, M. (1974), “A Conditional Logit Model of Occupational Choice”. Journal of Political Economy, 82, págs.389-398.

Cragg, J. y R. Uhler (1970), “The Demand for Automobiles”. Canadian Journal of Economics, 3, págs. 386-406.

Davidson, R. y J. MacKinnon (1984), “Several Tests for Model Specification In the Presence of Alternative Hypothesis”. Econometrica, 49, págs. 781-793.

Davidson, R. y J. MacKinnon (1993) Estimation and inference in Econometrics. Oxford University Press.

Dhrymes, P. (1984), "Limited Dependent Variables". En Griliches, Z. y M. Intriligator: Handbook of Econometrics, vol. 2, North Holland.

Goldberger, A. S. (2001), Introducción a la Econometría. Ariel.

Greene, W.H. (1997), Econometric Analysis, 3ª ed.. Prentice Hall.

Harvey, A. (1976), “Estimating Regression Models with Multiplicative Heteroscedasticity”. Econometrica, 44, págs. 461-465.

Hausman, J. (1978), “Specification Tests in Econometrics”. Econometrica, 46, págs.1251-1271.

Hausman, J. y D. McFadden (1984), “A Specification Test for the Multinomial Logit Model”. Econometrica, 52, págs. 1219-1240.

Hensher, D. (1986), “Simultaneous Estimation of Hierarchical Logit Mode Choice Models”. W.P. 24, MacQuarie University School of Economic and Financial Studies.

Luce, D. (1959), Individual Choice Behaviour. Wiley.

Maddala, G. S. (1983), Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press.

Manski, C. F. y D. McFadden (ed.)(1981), Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

McFadden, D. (1973). “Conditional Logit Analysis of Qualitative Choice Behaviour”. En P. Zarembka (ed.), Frontiers in Econometrics. Academic Press. Nueva York.

McFadden, D. (1974), “The Measurement of Urban Travel Demand”. Journal of Public Economics, 3, págs. 303-328.

McFadden, D. (1984), "Econometric Analysis of Qualitative Response Models". En Griliches, Z. y M. Intriligator: Handbook of Econometrics, vol. 2, North Holland.

Nelder, J.A. y R.W.M. Wedderburn (1972), “Generalized Linear Models” Journal of the Royal Statistical Society, A, 135, págs. 370-384.

Schmidt, P. y R. Strauss (1975a), “The Prediction of Occupation Using Multiple Logit Models”. International Economic Review, 16, págs. 471-486.

Schmidt, P. y R. Strauss (1975b), “Estimation of Models with Jointly Dependent Qualitative Variables. A Simultaneous Logit Approach”. Econometrica, 43, págs. 745-755.

Terza, J. (1985), “Ordinal Probit: A Generalization”. Communications in Statistics, 14, págs. 1-12.

Thurstone, L. (1927), “A Law of Comparative Judgement”. Psychological Review, 34, págs. 273-286.

Yatchew, A. y Z. Griliches (1984), “Specification Error in Probit Models”. Review of Economics and Statistics, 66, págs. 134-139.

Zarembka, P.(ed.) (1973), Frontiers in Econometrics, Academic Press.

Zavoina, R. Y W. McElvey (1975), “A Statistical Model for the Analysis of Ordinal Level Dependent Variables”. Journal of Mathematical Sociology, Verano 1975, págs. 103-120.

ENSAYO 3. ¿Influye el tamaño de una empresa sobre la probabilidad de acceder a la financiación bancaria?¿Cambia el sentido de la influencia a partir de un tamaño determinado?

Durante la realización de este ensayo el alumno tratará de encontrar una respuesta a la cuestión planteada y, en el camino, deberá ejercitarse en la utilización y comprensión de los **modelos de elección discreta logit y probit.** Como se sabe estos modelos son ampliamente utilizados en los estudios en los que la variable objetivo, en este ensayo la financiación bancaria de proyectos empresariales, facilita información sobre la toma de una decisión.

Los datos estadísticos necesarios para la parte empírica del ensayo proceden de una muestra representativa de empresas extraída de la *Encuesta de estrategias empresariales* (**EEE**) correspondiente al año 1994. Esta información se encuentra en el fichero *ensayo3.wf1*.

1. *Modelo teórico.*

Como se sabe, las empresas adoptan decisiones de financiación en condiciones de incertidumbre. Los contratos de financiación se caracterizan por la existencia de información asimétrica y la presencia de azar moral. En lo que sigue, trabajaremos con el supuesto de que tanto las empresas, que buscan financiación, como los agentes que proporcionan dicha financiación, financiación directa o intermediada, son neutrales ante el riesgo.

Las múltiples fuentes de financiación a las que una empresa acude pueden reducirse a un esquema con tres elementos: autofinanciación; financiación directa; y financiación intermediada. En este bosquejo simplificador la toma de una decisión de financiación cabe representarla como el resultado de un proceso de elección entre múltiples alternativas. En este ensayo, empero, centraremos exclusivamente nuestra atención en dos posibilidades: que la empresa acceda o que no acceda a la financiación bancaria. La reducción de las tres alternativas del esquema inicial a las dos que consideramos en este ensayo genera complicaciones de interpretación que deben ser tratadas con cuidado. En efecto, una empresa puede no acudir a una financiación bancaria por dos motivos diferentes: bien porque su acceso al crédito bancario esté cerrado, bien porque recurra a sus fondos propios o porque su solvencia y garantías le permitan acceder a una financiación directa.

Desde el punto de vista de los bancos, la presencia de información asimétrica en los contratos financieros hace que los primeros busquen tanto un sistema de garantías antes de conceder la financiación como que evalúen los costes de supervisión en los que incurrirán como consecuencia de la vigilancia del contrato. La influencia de las variables empresariales sobre la probabilidad de acceder al crédito bancario deberá interpretarse conforme al esquema anterior. Así, por ejemplo, el volumen elevado de flujos de caja influirá positivamente sobre la probabilidad de autofinanciación y, por ende, tendrá un efecto negativo sobre el acceso al crédito bancario. También el tamaño de la empresa será tanto un motivo de garantía para la concesión de un crédito como una señal de reducción de los costes de supervisión (en presencia de rendimientos crecientes de escala) y, en consecuencia, tendrá un influjo positivo sobre la probabilidad de acceder al crédito bancario. El sentido del efecto de otras variables determinantes ya ha sido considerado en clase.

Para una discusión completa de los aspectos teóricos de este ensayo puede verse

Suarez(2001).

2. Descripción de los datos.

Los datos que contiene el archivo *ensayo3.wf1* son desconocidos, ya que no han sido utilizados en las sesiones desarrolladas en el aula. Un examen detallado de los mismos, sin embargo, no sería de gran utilidad en este texto y se deja para el trabajo en el aula de informática.

Lo que sí resulta conveniente en este momento es el comentar ciertos aspectos relacionados con la transformación de ciertas variables, cuando esta transformación resulte útil para el trabajo empírico. De las variables que van a formar parte más adelante de nuestra especificación inicial una de las más relevantes es la representativa de los flujos de caja de la empresa. Aunque la información de esta variable se recoge directamente en la EEE, para evitar distorsiones causadas por el tamaño de la empresa se usa como índice el cociente entre los flujos de caja anuales y el volumen de ventas anual, la variable *caja* en el fichero *ensayo3.wf1*.

Como se ha señalado en el *apartado 1*, la influencia de la variable *caja* puede interpretarse en dos direcciones diferentes (pequeño y elevado volumen de flujos de caja) y va a tener presumiblemente un efecto no lineal sobre la probabilidad de acceder a un crédito bancario. Esta influencia no lineal la hemos representado en los ensayos realizados hasta ahora mediante términos cuadráticos de la propia variable y términos de interacción entre la variable considerada y otras variables relevantes. Esa representación es correcta y produce resultados razonables. Sólo tiene una dificultad empírica: obliga a que el efecto no lineal de una variable sobre la probabilidad tenga la forma de una parábola, con un máximo o un mínimo. Mediante esa especificación funcional no permitimos la presencia de máximos o mínimos locales. Esta representación no es suficientemente flexible en muchos casos.

Lo que interesa destacar aquí es que estos efectos no lineales cabe recogerlos de otra forma ampliamente utilizada en los estudios aplicados. Esta forma se denomina una función articulada. Para construir una función articulada se divide el rango (o recorrido) de la variable continua en un conjunto de tramos discretos. El número de tramos a considerar es un asunto aplicado y depende de la magnitud del rango y de los intervalos en los que se desee apreciar el comportamiento de la variable. Una vez definido el número y magnitud de los tramos se construyen tantas variables ficticias como tramos considerados. Estas variables ficticias toman el valor uno en el intervalo de definición del tramo y cero en el resto de los casos. La función del **EViews** útil para realizar este proceso es la **@recode** (ver *Ensayo 2* para una explicación de su funcionamiento en referencia a la variable *nivest*).

Apliquemos el procedimiento anterior a la variable *caja*. Esta variable tiene un rango que va de un valor mínimo de 0.0234 a un valor máximo de 106.53. En su histograma cabe apreciar que a partir del valor 40 las observaciones son muy escasas. Entre las muchas opciones de construcción de una función articulada (no hay sólo una), una que es razonable consiste en dividir el rango en cinco tramos con la magnitud siguiente: 0-10; 10-20; 20-30; 30-40; 40-máximo. Las variables ficticias correspondientes se definirán como:

genr `caja01=@recode(caja<=10,1,0)`
genr `caja02=@recode(caja>10 and caja<=20,1,0)`
...
Este proceso se reitera hasta el último tramo considerado caja05.

De esta forma, concluimos que la posible no-linealidad de la influencia de *caja* sobre la probabilidad de acceder al crédito bancario puede alternativamente recogerse mediante una función cuadrática de *caja* o mediante la inclusión de todas las variables ficticias anteriores menos una. Los términos de interacción deben especificarse en ambos casos. En la estimación, cada parámetro de las variables ficticias representará, como se sabe de otros ensayos, el efecto diferencial del volumen del flujo de caja comprendido en el tramo correspondiente a dicho parámetro, con respecto al volumen del flujo de caja recogido en la constante (el tramo no incluido en la especificación explícita del modelo).

Los gráficos individuales y por parejas de las variables que se utilizarán en este ensayo se dejan para el trabajo en clase de informática.

3. *Estimación de una ecuación.*

En este ensayo la variable objetivo, *bancos*, contiene información discreta. Toma el valor 1 en el caso en que la empresa considerada haya accedido en 1994 al crédito bancario; toma el valor 0 en el caso en que la empresa considerada no haya accedido en 1994 al crédito bancario.

Las variables determinantes de la probabilidad que se usarán en este ensayo son: *tamano*, tamaño de la empresa medido por el número de trabajadores de la misma; *caja*, flujo de caja dividido por las ventas; *sector*, sector de producción al que pertenece la empresa. En la zona de trabajo *ensayo3.wf1* hay otras variables relevantes para la explicación del acceso al crédito bancario. No van a ser incluidas, sin embargo, en nuestra ecuación. La explicación de por qué no se consideran en este ensayo viene de la escasez de variables disponibles en la zona de trabajo. Como se sabe, esta zona de trabajo también será usada por los asistentes al curso para la redacción de su propio ensayo sobre los determinantes de las fuentes de financiación de una empresa. Si se hubieran incluido en este ensayo todas las variables relevantes el resultado hubiera sido de *tierra quemada*, es decir, hubiera sido casi imposible que un asistente hubiera producido su propio ensayo sobre una información ya trabajada al detalle. Es por este motivo, y no por ningún otro, que el modelo a especificar se ha sometido a un proceso de autoreducción.

Construimos un modelo empírico que determine la probabilidad de tomar la decisión de acudir o no acudir al crédito bancario, la variable endógena para nuestros propósitos. Lo hacemos por medio de un modelo logit o probit. El argumento de la función de distribución, logística o normal, respectivamente, es una combinación lineal (en los parámetros) de un conjunto de variables débilmente exógenas. Como se sabe, para nuestros objetivos de aprendizaje no nos centraremos de una forma importante en la crítica del modelo empírico y partiremos del supuesto de que la exogeneidad débil de las variables determinantes se cumple.

3.1. Forma funcional.

Ya se ha comentado que el índice en nuestra ecuación logit o probit es una combinación lineal de parámetros y variables débilmente exógenas. Las posibles no linealidades en el efecto de las variables sobre la probabilidad de comprar un vehículo se recogerán en el índice por las dos vías comentadas en el *apartado 2*: formas cuadráticas y formas articuladas. Ambas especificaciones se trabajarán en clase de informática. En el texto se usa la especificación articulada.

3.2. Estrategia de modelización.

Se empleará la estrategia de lo general a lo particular, con una especificación general conducida principalmente por los criterios que la teoría económica marca. Los principales comentarios relacionados con las predicciones teóricas del modelo ya se han expuesto en el *apartado 1*.

Como modelo general la ecuación a estimar es:

$$\begin{aligned} \text{Prob(Bancos=1)} = {} & \alpha+\beta_1 \text{ tamano} + \beta_2 \text{ tamano\textasciicircum 2} + \beta_3 \text{ caja01} + \beta_4 \text{ caja02} + \beta_5 \text{ caja03} + \\ & \beta_6 \text{ caja04} + \beta_7 \text{ tamano*caja01} + \beta_8 \text{ tamano*caja02} + \\ & \beta_9 \text{ tamano*caja03} + \beta_{10} \text{ tamano*caja04} + \beta_{11} \text{ s1} + \ldots\ldots + \\ & \beta_{28} \text{ s18} + u \end{aligned} \qquad [9]$$

3.3. Método de estimación.

Con una variable objetivo discreta, los modelos logit y probit se estiman por el método de la máxima verosimilitud (MV).

La función de verosimilitud está formada por el producto de las contribuciones de cada una de las observaciones de la muestra (bajo el supuesto de independencia). Las observaciones para las que la variable *bancos* toma el valor 1 contribuyen con una probabilidad igual al valor de la función de distribución y las observaciones en que *bancos* toma el valor 0 contribuyen con el complementario de la función de distribución. Como es usual, la función que realmente se maximiza es la función de verosimilitud en logaritmos. Su expresión está recogida en la fórmula (5) del *Ensayo 2*.

4. Análisis de resultados.

Para estimar la ecuación en (9) deberemos, en primer lugar, generar las variables que se encuentran en la parte derecha de la ecuación relativas a los tramos de la función articulada de la variable *caja* y las representativas de los distintos sectores económicos.

Luego estimamos el modelo con la opción **BINARY-binary choice (logit, probit, extreme value).** El resultado del proceso para el modelo probit se encuentra en el *Cuadro 11* de la página siguiente. El resultado de la estimación de un logit con idéntica especificación se da por reproducido y se realizará en el aula de informática.

Para analizar los resultados anteriores seguiremos el esquema planteado en clase y desarrollado en el *Ensayo 1*.

4.1. Contrastes.

Como se sabe, en cualquier modelo estimado se deben realizar contrastes de falta de especificación (especificación errónea) y de especificación a fin de verificar la validez del mismo. Seguiremos el orden ya establecido en el *Ensayo 1*.

Cuadro 11. Resultados de la estimación de un probit de la ecuación en (9): Modelo general.

```
============================================================
Dependent Variable: BANCOS
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 04/22/02   Time: 15:22
Sample: 1 1027
Included observations: 1027
Convergence achieved after 9 iterations
Covariance matrix computed using second derivatives
============================================================
     Variable      CoefficientStd. Errorz-Statistic  Prob.
============================================================
      TAMANO        -9.49E-05   0.001778  -0.053387   0.9574
     TAMANO^2       -3.96E-07   1.29E-07  -3.075484   0.0021
      CAJA01         0.172485   0.322483   0.534866   0.5927
      CAJA02         0.184241   0.325439   0.566132   0.5713
      CAJA03        -0.088745   0.348997  -0.254285   0.7993
      CAJA04        -0.167521   0.411686  -0.406914   0.6841
   CAJA01*TAMANO     0.002424   0.001809   1.339764   0.1803
   CAJA02*TAMANO     0.001407   0.001798   0.782555   0.4339
   CAJA03*TAMANO     0.001783   0.001850   0.964047   0.3350
   CAJA04*TAMANO     0.000603   0.001945   0.310003   0.7566
        S1          -0.481779   0.276925  -1.739746   0.0819
        S10         -0.149270   0.248276  -0.601229   0.5477
        S11         -0.086342   0.224937  -0.383849   0.7011
        S12         -0.380579   0.305957  -1.243896   0.2135
        S13          0.113775   0.211916   0.536888   0.5913
        S14         -0.223776   0.222561  -1.005460   0.3147
        S15         -0.125554   0.352617  -0.356064   0.7218
        S16         -0.050054   0.215837  -0.231907   0.8166
        S17         -0.707841   0.284063  -2.491851   0.0127
        S18         -0.455906   0.403746  -1.129191   0.2588
        S2          -0.158189   0.208936  -0.757116   0.4490
        S3          -0.171879   0.302480  -0.568234   0.5699
        S4          -0.063720   0.200228  -0.318236   0.7503
        S5          -0.093056   0.262924  -0.353928   0.7234
        S6           0.399616   0.363544   1.099222   0.2717
        S7           0.847292   0.406803   2.082805   0.0373
        S8          -0.123977   0.242814  -0.510586   0.6096
        S9          -0.210805   0.215836  -0.976691   0.3287
         C           0.194234   0.333150   0.583025   0.5599
============================================================
Mean dependent var   0.674781    S.D. dependent var 0.468685
S.E. of regression   0.454804    Akaike info criteri1.226755
Sum squared resid    206.4329    Schwarz criterion  1.366091
Log likelihood      -600.9389    Hannan-Quinn criter1.279644
Restr. log likelihoo-647.7711    Avg. log likelihoo-0.585140
LR statistic (28 df) 93.66431    McFadden R-squared 0.072297
Probability(LR stat) 5.25E-09
============================================================
Obs with Dep=0            334     Total obs            1027
Obs with Dep=1            693
======================================================================
```

4.1.1. Contrastes de falta de especificación.

Para analizar el comportamiento de los residuos del modelo logit (o probit) se debe antes reflexionar sobre el carácter de dichos residuos. Hay que decir una vez más que los residuos que presentan propiedades análogas a las de los residuos MC se denominan residuos generalizados y se obtienen a partir de las ecuaciones de verosimilitud. Su expresión se encuentra en la fórmula (7) del *Ensayo 2*. También en dicho ensayo se encuentra la forma de obtenerlos en el **EViews**.

Los contrastes de especificación errónea que se suelen establecer en estos modelos son los de variables omitidas y los de heteroscedasticidad, además de los correspondientes gráficos de residuos. El tipo de contrastes a realizar no tiene nada que ver con los análogos del modelo de regresión con variables del lado izquierdo continuas.

Realizaremos un contraste de heteroscedasticidad en nuestro modelo general. El contraste de heteroscedasticidad descrito en Davidson y Mackinnon (1993) es un contraste de multiplicadores de Lagrange (LM) diseñado para modelos probit. Como ya se experimentó en la realización del *Ensayo 2*, la puesta en práctica del contraste instrucción a instrucción (lo que los investigadores de economía aplicada denominan *método del manubrio*) es una tarea tediosa. Para facilitar en este ensayo su desarrollo se ha diseñado un programa denominado *dmck01.prg.* La explicación de cómo usar un programa de **Eviews** para acortar el tiempo empleado en determinados menesteres engorrosos se hará en el aula informática. Detalles más concretos sobre la forma del contraste y su realización por el método del *manubrio* se encuentran en el *Ensayo 2*.

En su versión actual, el programa *dmck01.prg* exige el siguiente prerrequisito. Antes de ejecutarse debe construirse en la zona de trabajo un grupo formado por las variables del lado derecho de la ecuación en (9), con la excepción del término independiente. Llamamos al grupo *probit_x*. Para ello, en la ventana superior del **EViews** escribimos **group probit_x** y, a continuación, relacionamos el nombre de las variables del lado derecho de (9). Como se sabe, el procedimiento mejor para escribir las variables es copiar el texto en la ventana de **estimate equation** y pegarlo en la ventana superior.

Una vez hecho lo anterior, para la ejecución del programa resulta conveniente (no es necesario) cargar el programa en la ventana de trabajo. Si se hace así, el programa aparece sobre la pantalla y se puede leer su contenido, en el caso de que fuere necesario. Para cargar el programa *pinchamos* **file/open/program/dmck01.prg**. La ejecución del programa consiste en *pinchar* **run** y escribir en la ventana que aparece en blanco el nombre de la variable del lado izquierdo de la ecuación en (9), en primer lugar, y el de la que causa presumiblemente la heteroscedasticidad, en segundo lugar, y en este orden estricto. Esa secuencia basta para que el programa suministre el resultado del contraste: el valor del estadístico de multiplicadores de Lagrange (LM), variable *lm_test*, y el valor de la probabilidad asociada a dicho estadístico, variable *p_value*. Debe recordarse del *Ensayo 2* que las dos variables anteriores son escalares y, por tanto, para visualizar su valor, se deberá hacer doble *click* sobre su nombre (el valor aparecerá en la zona inferior izquierda de la ventana del **EViews**).

El resultado de poner en marcha el programa *dmck01.prg* para la ecuación en (9) es el contenido en el *Cuadro 12*, siguiente.

Cuadro 12. Resultados del contraste de heteroscedasticidad en (9) para diferentes alternativas de la variable causante de la heteroscedasticidad.

Variable causante de la heteroscedasticidad	Estadístico LM	Probabilidad
Tamano	1.169	0.2796
Tamano^2	0.014	0.9043
$\text{Tamano}^{1/2}$	1.571	0.2100
Caja	0.9292	0.3351
Caja^2	0.8617	0.3533
$\text{Caja}^{1/2}$	0.7866	0.3751

Como puede verse en el *Cuadro 12*, el contraste no rechaza la hipótesis nula en ninguna de las posibilidades contempladas. Aunque se pudieran incluir otras opciones, no se realizan en este momento para evitar la situación de *tierra quemada* antes comentada.

Cabe concluir en este momento que la especificación del modelo al menos supera el contraste de heteroscedasticidad propuesto para modelos probit por Davidson y MacKinnon. Como ya se comentó en los *Ensayos 1, y 2*, en ejercicios aplicados es muy difícil que todos los contrastes de especificación sean superados por los modelos especificados. Por ello debemos mostrarnos satisfechos de que al menos el modelo no sea heteroscedástico.

4.1.2. Contrastes de especificación.

La utilización de contrastes de especificación individuales, estadísticos t, y contrastes conjuntos, estadísticos F, contrastes de Wald (lineales y no lineales) y de la razón de verosimilitud (o verosimilitudes) (LR), ha sido abundante en los ensayos precedentes. Se aprovechará esta ocasión para dar indicaciones sobre el uso de estos contrastes que proceden de la experiencia del trabajo aplicado.

Como se ve en el *Cuadro 11,* las variables del lado derecho de la ecuación pueden agruparse en cuatro colectivos distintos. El primero formado por la variable *tamano* y su cuadrado. El segundo constituido por las variables ficticias que integran la función articulada que representa a la variable *caja*. El tercero integrado por los términos de interacción entre las variables *tamano* y *caja*. El cuarto compuesto por las variables ficticias que simbolizan el efecto de los sectores de actividad de la empresa. De estos cuatro colectivos hay tres que tienen contenido discreto. Es decir, la función articulada, los términos de interacción y los sectores económicos contienen datos discretos (0 y 1, en algunos casos, y 0 y valores mayores que cero, en otros) y recogen los efectos diferenciales entre la variable considerada y la incluida en la constante del modelo. Por esta razón, si deseamos analizar si son significativos los efectos de una variable concreta sobre la probabilidad de acceder a la financiación bancaria no tiene sentido evaluar la significatividad estadística individual del parámetro de una variable concreta, dentro de un colectivo determinado, sino analizar la significatividad conjunta de todos los parámetros de las variables que integran el colectivo.

Por ejemplo, si queremos saber si la contribución de los factores específicos ligados a la pertenencia a un sector económico concreto afectan significativamente a la probabilidad de acceso al crédito bancario, no tiene sentido analizar la significatividad individual de la variable *s7*, que simboliza que la empresa se encuentra en el sector de *Material y accesorios eléctricos*, ya que la significatividad de su parámetro correspondiente (t=2.08 con Prob=0.037, por lo tanto significativo) sólo nos indica que el comportamiento diferencial de ese sector con el incluido en la constante es distinto de cero. Como la inclusión de un sector concreto en la constante es arbitraria, el resultado anterior no nos dice nada sobre la relevancia de los sectores económicos en la explicación de la probabilidad del acceso al crédito bancario.

Lo correcto en el contraste planteado en el párrafo anterior es verificar la hipótesis nula de que todos los parámetros vinculados al colectivo de sectores económicos sean conjuntamente nulos. El rechazo de la hipótesis nula sería un argumento a favor de la relevancia de la pertenencia a un sector económico en la determinación de la probabilidad de acceso al crédito bancario. El no rechazo de la hipótesis nula indicaría que los factores específicos ligados a la pertenencia a un sector económico concreto no son relevantes para dicha probabilidad.

El mismo argumento desarrollado en los párrafos anteriores se puede aplicar a los colectivos de la función articulada de la variable *caja* y a los términos de interacción del modelo.

Debe recordarse, no obstante, que el principal argumento a utilizar en la especificación de una ecuación discreta es la teoría económica apropiada, por lo que las razones estadísticas deben ser acordes a las predicciones del modelo económico.

En nuestro caso, los contrastes realizados han sido del tipo LR con el fin de investigar la dirección de una posible estrategia de simplificación de la ecuación (9). La tarea concreta se reserva para el trabajo en el aula de informática pero los resultados logrados se resumen en el *Cuadro 13*, siguiente.

Cuadro 13. Resultados de los contrastes de especificación. Contrastes LR de nulidad conjunta de colectivos concretos de variables en la ecuación (9).

Colectivos de variables redundantes	Estadístico F	Probabilidad	Estadístico LR	Probabilidad
Sector	1.336	0.1568	25.88	0.1025
Sector y términos de interacción	1.499	0.06513	36.25	0.02860
Sector, términos de interacción y función articulada	1.744	0.01216	49.80	0.003324

Como se ve en el *Cuadro 13*, la simplificación de los tres colectivos es rechazada por los datos. En lo que respecta a la simplificación de los colectivos de sector y términos de interacción, hay que decir, como es habitual, que el estadístico F no la rechaza y el estadístico LR sí. Por motivaciones asintóticas y por razones de conveniencia interpretativa utilizaremos la información del estadístico LR y únicamente simplificaremos los términos que representan los distintos sectores económicos. Para una mayor profundización en el tema habría que investigar cualitativamente el carácter de los sectores económicos considerados, tarea que escapa a los objetivos de este ensayo.

4.2. Bondad del ajuste y medidas de información.

Las características principales de las medidas de la bondad del ajuste y de información que aparecen en las salidas de las estimaciones del **EViews** se trataron en el *Ensayo 2*.

En este ensayo se comentarán más adelante a la hora de comparar las estimaciones del modelo general y del simplificado.

4.3. Interpretación económica y verificación de las predicciones iniciales.

El proceso de modelización de lo general a lo particular, en la dirección señalada anteriormente en el epígrafe 4.1, conduce a la especificación siguiente:

$$\begin{aligned}\text{Prob(Bancos=1)} = {} & \alpha+\beta_1 \text{ tamano} + \beta_2 \text{ tamano\^{}2} + \beta_3 \text{ caja01} + \beta_4 \text{ caja02} + \beta_5 \text{ caja03} + \\ & \beta_6 \text{ caja04} + \beta_7 \text{ tamano*caja01} + \beta_8 \text{ tamano*caja02} + \\ & \beta_9 \text{ tamano*caja03} + \beta_{10} \text{ tamano*caja04} + u \qquad [10]\end{aligned}$$

En esta dirección de simplificación cualquier otra restricción de nulidad adicional queda rechazada por los datos, o resulta incoherente con las predicciones de la teoría económica. Por tanto, se acepta en principio esta especificación en (10) como el modelo más simple que pueda representar la cuestión contemplada. A continuación, a fin de confirmar la validez del modelo volvemos a realizar los contrastes de especificación errónea. En concreto el contraste de heteroscedasticidad. Finalmente, analizamos en términos comparativos, modelos general y simplificado, las medidas de ajuste e información. La estimación de la ecuación en (10) es la que se encuentra en el *Cuadro 14*, siguiente.

Cuadro 14. Resultado de la estimación de la ecuación en (10): Modelo simplificado.

```
============================================================
Dependent Variable: BANCOS
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 04/24/02   Time: 18:10
Sample: 1 1025
Included observations: 1025
Convergence achieved after 10 iterations
Covariance matrix computed using second derivatives
============================================================
     Variable      CoefficientStd. Errorz-Statistic  Prob.
============================================================
      TAMANO        -0.000311   0.001765  -0.175953   0.8603
     TAMANO^2       -3.61E-07   1.15E-07  -3.143929   0.0017
      CAJA01         0.131027   0.316736   0.413678   0.6791
      CAJA02         0.147403   0.319513   0.461337   0.6446
      CAJA03        -0.138974   0.343345  -0.404765   0.6857
      CAJA04        -0.236689   0.405882  -0.583146   0.5598
   CAJA01*TAMANO     0.002343   0.001794   1.306000   0.1916
   CAJA02*TAMANO     0.001453   0.001785   0.814102   0.4156
   CAJA03*TAMANO     0.001772   0.001827   0.969659   0.3322
   CAJA04*TAMANO     0.000774   0.001928   0.401597   0.6880
         C           0.152612   0.308396   0.494859   0.6207
============================================================
Mean dependent var   0.676098    S.D. dependent var 0.468192
S.E. of regression   0.456604    Akaike info criteri1.219278
Sum squared resid    211.4059    Schwarz criterion  1.272212
Log likelihood      -613.8800    Hannan-Quinn criter1.239372
Restr. log likelihoo-645.5205    Avg. log likelihoo-0.598907
LR statistic (10 df) 63.28088    McFadden R-squared 0.049015
Probability(LR stat) 8.63E-10
============================================================
Obs with Dep=0            332     Total obs            1025
Obs with Dep=1            693
============================================================
```

Realizado un contraste de heteroscedasticidad sobre la estimación del *Cuadro 14*, resulta que la hipótesis nula no es rechazada. Los valores del estadístico LM y su probabilidad asociada son, respectivamente, 0.4340 y 0.5100, para el caso de que la variable *tamano* causara la heteroscedasticidad y 1.379 y 0.2402, para el caso de que fuera la variable *caja*. La verificación de otras alternativas se deja para la clase de informática.

Tampoco hay diferencias notables en los valores de las medidas de ajuste y de información. Si bien el valor del R^2 de McFadden es inferior en el modelo simplificado que en el general, los criterios de Akaike, Schwarz y Hannan-Quinn proporcionan valores inferiores en el modelo simplificado a los obtenidos en el modelo general.

Cabe concluir que dado que en el proceso de simplificación las propiedades estadísticas no se han deteriorado y las medidas de información son similares en el modelo simplificado y el modelo general, el modelo simplificado o reducido es preferible al modelo general por el principio de parsimonia.

A partir de este momento, la interpretación económica seguirá dos direcciones diferentes: por un lado, se profundizará, con la ayuda de un gráfico, en el significado económico de la estimación del modelo final y se dará respuesta a la pregunta inicial; por otro, se usará el modelo simplificado para realizar predicciones sobre el comportamiento de las empresas ante la financiación bancaria. En ambas direcciones se validarán o rechazarán las predicciones de partida.

4.3.1. Significado económico y respuesta a la pregunta inicial.

La interpretación económica tendrá como objetivo la explicación de las diferencias que el tamaño de una empresa provoca sobre la probabilidad de que ésta acuda a una financiación bancaria para el desarrollo de sus proyectos. En este ensayo, la probabilidad es además función de los flujos de caja de la empresa, con interacciones entre ambas variables. El carácter de la presentación de la variable *caja* como función articulada influirá sobre el tipo de análisis, como veremos a continuación.

Escogeremos como forma de presentación un gráfico en el que en el eje de ordenadas estarán los valores de la probabilidad de acceder al crédito bancario y en el de abscisas el número de trabajadores de una empresa. Como la magnitud de los flujos de caja de la empresa se ha representado en forma articulada, en el gráfico se puede representar cómo se altera la relación entre probabilidad y tamaño para las empresas que se sitúan en un tramo concreto de sus flujos de caja, es decir, para aquellas empresas cuyos flujos de caja tiene un volumen concreto. La idea es simple, ya que existe interacción entre tamaño y flujos de caja, esa interrelación se trasladará a la forma de la relación probabilidad-tamaño. La forma de realizar este tipo de gráficos se ha descrito en los ensayos anteriores y se ha practicado ampliamente en el aula de informática, por lo que en esta redacción se omiten detalles innecesarios.

En concreto, construiremos un gráfico en el que se plasmarán los resultados de las ecuaciones siguientes:

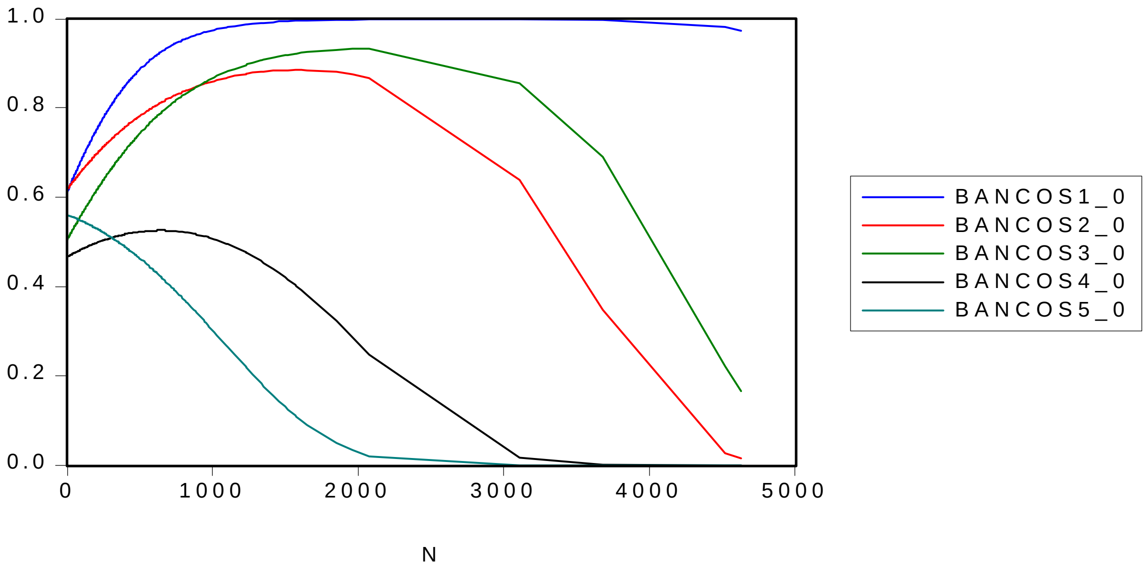
```
BANCOS1_0  = 1  - @CNORM(-(-0.0003105238172  * TAMANO  - 3.613323648e-07  * (TAMANO^2)  + 0.1310266762  + 0.002342822469  * TAMANO  + 0.1526123666))

BANCOS2_0 = 1  - @CNORM(-(-0.0003105238172  * TAMANO  - 3.613323648e-07  * (TAMANO^2)  + 0.1474033235  + 0.00145281096  * TAMANO  + 0.1526123666))

BANCOS3_0 = 1  - @CNORM(-(-0.0003105238172  * TAMANO  - 3.613323648e-07  * (TAMANO^2)  - 0.138973835  + 0.001771571205  * TAMANO  + 0.1526123666))

BANCOS4_0 = 1  - @CNORM(-(-0.0003105238172  * TAMANO  - 3.613323648e-07  * (TAMANO^2)  - 0.2366887409  + 0.0007743143636  * TAMANO  + 0.1526123666))

BANCOS5_0  = 1  - @CNORM(-(-0.0003105238172  * TAMANO  - 3.613323648e-07  * (TAMANO^2)  + 0.1526123666))
```
[11]

Las ecuaciones en (11) son las expresiones de la probabilidad de que una empresa acceda al crédito bancario para distintos tramos de su volumen de flujos de caja: 0-10 (*bancos1_0*); 10-20 (*bancos2_0*); 20-30 (*bancos3_0*); 30-40 (*bancos4_0*); 40 y más (*bancos5_0*). Como se ve la no linealidad se manifiesta en que para distintos tramos de la variable *caja* se altera también la pendiente del índice lineal relativa a la variable *tamano*.

El gráfico que puede ser interpretado es el *Gráfico 12* de la página siguiente. A primera vista se produce un resultado esperado: la probabilidad no es una función lineal del tamaño sino que su influencia presenta cambios de sentido en el rango de tamaños. Además, los resultados son muy diferentes si tenemos en cuenta el volumen de flujos de caja de la empresa. Empresas con elevado volumen de flujos de caja presentan una disminución uniforme de la probabilidad de acceder al crédito bancario a medida que su tamaño aumenta. No parece disparatado argumentar que el sentido de la relación se debe al incremento de la autofinanciación.

A medida que las posibilidades de autofinanciación desaparecen (menor volumen de flujos de caja) la probabilidad de acceso al crédito bancario adquiere una forma parabólica con máximos situados para tamaños entre dos mil y tres mil trabajadores.

Gráfico 12. Probabilidad de acceder al crédito bancario en términos del tamaño de la empresa. Efectos diferenciales debidos al volumen relativo de flujos de caja (con respecto al volumen de ventas).



Un último aspecto a destacar es el corte de las curvas en el gráfico que comentamos. Esos cortes indican que no sólo hay interacción entre las variables de tamaño y flujos de caja sino que sus efectos combinados alteran la ordenación de probabilidades para distintos tamaños de empresa. Otros comentarios se reservan para la sesión en el aula de informática.

4.3.2. Predicciones.

La comparación de las predicciones de nuestro modelo con las observaciones reales del comportamiento de las empresas es otro aspecto interesante del trabajo aplicado. En este caso hacemos referencia no a la predicción de una probabilidad sino a la predicción del comportamiento real de una empresa que reúne una serie de características concretas. Nuestro fin es la predicción de si la empresa accederá al crédito bancario (*bancos*=1) o no (*bancos*=0). El **EViews** proporciona una tabla de las predicciones para cualquier modelo discreto estimado. Como se sabe, la predicción implica la definición previa de un umbral a partir del cuál una observación se clasifica como 1 o como 0. Para muestras equilibradas el umbral razonable es el valor 0.5. En presencia de muestras desequilibradas la elección anterior es como poco dudosa.

Para elaborar una tabla de predicciones basta *pinchar* **view/expectation-prediction table.** Aparecerá una ventana en la que se pregunta por el valor del umbral a utilizar. Por defecto, figura en élla el valor 0.5. En el *Cuadro 15* de la página siguiente se encuentra el resultado de predecir con el modelo final en (10).

La interpretación de la salida del **EViews** en este caso es prolija. Su detalle se realizará en el aula de informática. Baste decir aquí que la búsqueda por parte de los programadores del **EViews** de suministrar una mayor información a los usuarios ha resultado en este caso en una complicación innecesaria de la habitual tabla de predicciones, que resume los aciertos y errores cometidos. Esa tabla tradicional ocupa en el *Cuadro 15* las seis primeras filas y las tres primeras columnas. Como puede verse, si fijamos en nuestro análisis el umbral en el valor 0.5, el acierto

Cuadro 15. Tabla de predicciones del modelo final del Cuadro 14.

```
============================================================
Dependent Variable: BANCOS
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 04/24/02   Time: 18:10
Sample: 1 1025
Included observations: 1025
Prediction Evaluation (success cutoff C = 0.5)
============================================================
                       Estimated Equ          Constant Probability
               Dep=0   Dep=1   Total   Dep=0   Dep=1   Total
============================================================
P(Dep=1)<=C       17       9      26       0       0       0
 P(Dep=1)>C      315     684     999     332     693    1025
   Total         332     693    1025     332     693    1025
  Correct         17     684     701       0     693     693
 % Correct      5.12   98.70   68.39    0.00  100.00   67.61
% Incorrect    94.88    1.30   31.61  100.00    0.00   32.39
Total Gain*     5.12   -1.30    0.78
Percent Gain**  5.12      NA    2.41
============================================================
                       Estimated Equ          Constant Probability
               Dep=0   Dep=1   Total   Dep=0   Dep=1   Total
============================================================
E(# of Dep=0) 120.19  212.12  332.30  107.54  224.46  332.00
E(# of Dep=1) 211.81  480.88  692.70  224.46  468.54  693.00
   Total      332.00  693.00 1025.00  332.00  693.00 1025.00
  Correct     120.19  480.88  601.07  107.54  468.54  576.07
 % Correct     36.20   69.39   58.64   32.39   67.61   56.20
% Incorrect    63.80   30.61   41.36   67.61   32.39   43.80
Total Gain*     3.81    1.78    2.44
Percent Gain**  5.64    5.50    5.57
============================================================
*Change in "% Correct" from default (constant probability) specificati
**Percent of incorrect (default) prediction corrected by equation
============================================================
```

es notable en la predicción de la empresas que acceden al crédito bancario y también es notable el error que se comete en la predicción de las empresas que no acceden al crédito bancario. Este es el resultado que habitualmente se obtiene cuando se trabaja con una muestra desequilibrada como la de este ensayo, en que el 67.61% de la misma son empresas que acceden al crédito bancario.

Ejercicios de predicción con el mismo modelo y otros valores del umbral se dejan para el trabajo en el aula de informática.

Bibliografía.

Suarez, C. (2001) *Análisis de la información asimétrica en las decisiones financieras y reales de las empresas.* Tesis Doctoral. Universidad de Alcalá.