CUADRO Núm. 16

COEFICIENTES DE CORRELACION Y DE DETERMINACION MULTIPLES

	AÑOS	R	R^2
Trigo	1733-47	0,90	0,81
	1762-86	0,82	0,64
Cebada	1733-47	0,91	0,81
	1762-86	0,80	0,64

NOTA: Debido a las exigencias de la revista en cuanto al número de folios de cada artículo, no ha sido posible incluir en este apéndice algunas series de precios de ventas de trigo y cebada de ciertas localidades, las cuales están a disposición de todo el que quiera consultarlas.

TRATAMIENTO ECONOMETRICO DE LAS VARIABLES CUALITATIVAS

Gérard Lassibille
Lucía Navarro Gómez

0. INTRODUCCION

En microeconomía muchos comportamientos son cualitativos, por lo que para explicarlos deberá ampliarse el modelo de regresión clásico, de manera que incluya también variables cualitativas (tales como educación, profesión, sexo, estado civil. . .), representadas como variables ficticias. Podría pensarse que estos fenómenos se resisten al tratamiento cuantitativo, sin embargo, el uso de los métodos clásicos de análisis, específicamente diseñados para variables cuantitativas, puede también extenderse para tratar con variables ficticias, que sirven para recoger diversos tipos de efectos (además de los cualitativos, temporales, espaciales, etc.) en la misma relación.

Las variables ficticias pueden ser binarias o múltiples. Las primeras son variables dicotómicas, es decir que sólo toman dos valores, cero o uno; las últimas son una generalización de las primeras. Tanto unas como otras pueden ser a su vez explicativas o explicadas. La estimación de un modelo que contenga sólo variables cualitativas independientes, puede estimarse por los métodos ordinarios sin dificultad. Sin embargo, la presencia de variables dependientes cualitativas en un modelo, origina problemas de inferencia estadística especiales, ya que en este caso, las hipótesis generalmente admitidas en los modelos, tanto relativas a las perturbaciones como a la forma de las relaciones que ligan a las variables, ya no se cumplen, con lo cual se hace inoperante el método clásico de estimación de los mínimos cuadrados ordinarios.

Este artículo trata de explicitar las soluciones dadas, en la literatura econométrica, a la esimación de un modelo cuya variable dependiente sea cualitativa.

1. LA FUNCION DE PROBABILIDAD LINEAL

Si queremos estimar la probabilidad de realización de un suceso E, dado un cierto número de características asociadas al mismo, bajo la hipótesis de una representación lineal del fenómeno, el modelo econométrico subyacente ya no cumple las hipótesis clásicas de los mínimos cuadrados ordinarios, ni tampoco la definición de probabilidad. Si el primer inconveniente puede paliarse conservando el modelo lineal y estimándolo por el método de los mínimos cuadrados generalizados, el segundo inconveniente subsiste aún en este caso y la única forma de subsanarlo consiste en la estimación de un modelo no lineal.

1.1. Consecuencia del carácter binario de la variable dependiente

Consideremos el suceso:

E = {Obtención de un título universitario por un individuo, mediante un examen de fin de carrera.}

Especificamos:

$y_i = 1$ Si se realiza el suceso para el individuo i

$y_i = 0$ en caso contrario

Supongamos que la variable y_i, que en este caso es una variable binaria, está determinada por un conjunto de k variables exógenas independientes y fijas, x_k (binarias o no). Bajo el supuesto de relación lineal tendremos el modelo:

$$\text{(I)} \quad y_i = \beta_0 + \beta_1 x_{1i} + \ldots\ldots\ldots + \beta_k x_{ki} + \epsilon_i; \; i = 1, \ldots, n$$

donde ϵ_i es un término de perturbación aleatoria aditivo.

El modelo podemos también escribirlo así:

$$y_i = x_i \beta + \epsilon_i \qquad i = 1, \ldots, n$$

donde β es el vector de orden (k+1, 1) de los parámetros desconocidos y x_i el vector de orden (1, k+1) de las variables explicativas asociadas al individuo i.

La variable endógena y_i, al ser una función de la perturbación aleatoria, es también una variable aleatoria, que sólo puede tomar dos valores en este caso; por esta razón, su esperanza matemática condicional es igual a:

$$E(y_i \mid x_i) = \text{Prob}\,(y_i = 0 \mid x_i) \,.\, 0 + \text{Prob}\,(y_i = 1 \mid x_i) \,.\, 1 = \\ = \text{Prob}\,(y_i = 1 \mid x_i)$$

Bajo la hipótesis de que la esperanza de las perturbaciones ϵ_i es nula, tenemos:

$$E(y_i \mid x_i) = x_i \beta$$

con lo cual

$$\text{Prob}\,(y_i = 1 \mid x_i) = x_i \beta$$

El carácter dicotómico de la variable endógena hace que su esperanza matemática sea la probabilidad condicional de realización del suceso explicado, dado el vector de variables exógenas x_i (De aquí que se haya dado el nombre de función de probabilidad lineal al modelo anterior). Es decir, que el valor calculado del modelo es una estimación de la probabilidad condicional de realización asociada al valor 1 de la variable endógena.

Este modelo no satisface las hipótesis para la estimación por el método de los MCO. En efecto, calculemos para ello la variancia del término de perturbación. Este no puede tomar más que dos valores:

$$\epsilon_i = 1 - x_i \beta \qquad \text{si} \quad y_i = 1$$
$$\epsilon_i = - \; x_i \beta \qquad \text{si} \quad y_i = 0$$

Admitiendo la hipótesis de que la esperanza de ϵ es cero, tenemos:

$$E(\epsilon_i) = \text{Prob}\ (\epsilon_i = -x_i \beta)\ (-x_i \beta) + \text{Prob}\ (\epsilon_i = 1 - x_i \beta)\ (1 - x_i \beta) = 0$$

y como

$$\text{Prob}\ (\epsilon_i = -x_i \beta) + \text{Prob}\ (\epsilon_i = 1 - x_i \beta) = 1$$

tendremos:

$$\text{Prob}\ (\epsilon_i = -x_i \beta) = 1 - x_i \beta$$

y

$$\text{Prob}\ (\epsilon_i = 1 - x_i \beta) = x_i \beta$$

La variancia de la perturbación aleatoria se expresa en el caso discreto de la forma siguiente:

$$\text{Var}\ (\epsilon_i) = E(\epsilon_i^2) = \text{Prob}\ (\epsilon_i = -x_i \beta)\ (-x_i \beta)^2 + \text{Prob}\ (\epsilon_i = 1 - x_i \beta)\ (1 - x_i \beta)^2$$

y reemplazando las probabilidades por sus valores:

$$E(\epsilon_i^2) = x_i \beta\ (1 - x_i \beta) = \sigma_{ii}^2$$

y

$$E(\epsilon_j^2) = x_j \beta\ (1 - x_j \beta) = \sigma_{jj}^2$$

Luego, al no ser homoscedásticas las perturbaciones, los estimadores de los parámetros desconocidos del modelo, obtenidos por el método de los MCO, aunque sean lineales e insesgados son, sin embargo, ineficientes, es decir no tienen variancia mínima. Olvidar la heterocedasticidad de las perturbaciones significaría subestimar las verdaderas variancias de los estimadores y, consecuentemente, sesgar los tests hacia la aceptación de hipótesis.

1.2. Una solución a la estimación de un modelo con variable dependiente binaria: El método de los MCG

Para paliar el problema de la heterocedasticidad de las perturbaciones, puede estimarse el modelo por el método de los MCG o método de Aitken. Este consiste en transformar el modelo inicial de manera que ex-post las perturbaciones se conviertan en homocedásticas, para ello basta con ponderar cada observación por la inversa de la desviación estándar de la perturbación correspondiente.

En nuestro modelo (I) con

$$E(\epsilon_i) = 0 \quad \forall i$$

y

$$E(\epsilon_i \epsilon_j) = \sigma_{ii}^2 = \sigma^2 \lambda_i^2 \qquad \text{para } i = j$$
$$= 0 \qquad \text{para } i \neq j$$

Si dividimos por la desviación estándar de la perturbación:

$$\frac{y_i}{\lambda_i} = \frac{\beta_0}{\lambda_i} + \beta_1 \frac{x_{i1}}{\lambda_i} + \ldots + \beta_k \frac{x_{ki}}{\lambda_i} + \frac{\epsilon_i}{\lambda_i}$$

$$i = 1, \ldots, n$$

La perturbación transformada: $\frac{\epsilon_i}{\lambda_i}$ es ya homocedástica, como puede comprobarse fácilmente.

En estas condiciones puede aplicarse al modelo transformado el método de los MCO, el cual permitirá encontrar los estimadores minimocuadráticos generalizados del modelo inicial. Estos estimadores lineales e insesgados, tienen variancia mínima, siendo en consecuencia estimadores óptimos (ELIO).

Sin embargo, en el caso que estamos estudiando, las variancias de las perturbaciones son desconocidas y entonces el método puro de Aitken no es realizable. Para ello sería necesario dar antes una estimación consistente de cada perturbación. Mac Gillivray (1) sugiere tomar como estimador de la variancia la siguiente expresión:

$$\widehat{\text{Var}}\,(\epsilon_i) = \hat{y}_i\,(1-\hat{y}_i)$$

donde $\hat{y}_i$ es el valor calculado del modelo

$$y_i = x_i\,\beta + \epsilon_i$$

estimado por el método de los MCO.

Como no está excluido que haya variancias negativas, una posible solución para evitarlo sería elegir como expresión del estimador de la variancia la siguiente:

$$\widehat{\text{Var}}\,(\epsilon_i) = |\ \hat{y}_i - (1-\hat{y}_i)\ |$$

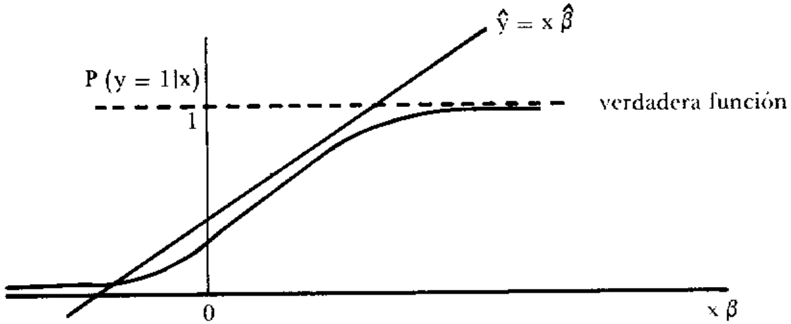
con lo cual estaría resuelto el problema.

Ahora bien, aunque la estimación de la función de probabilidad lineal por el método de los MCG constituye una mejora, en relación a su estimación por los MCO, sin embargo un problema importante subsiste aún: se trata de que la hipótesis generalmente

(1) Mac Gillivray en *Económetrica*, vol. 38, núm. 5, 1970. Págs. 775-776.

admitida de normalidad de las perturbaciones, postulada para juzgar la significación de las variables, no es aceptable en el caso de un modelo con variable dependiente binaria. En efecto, la perturbación aleatoria toma, en este caso, sus valores en el intervalo $(-x_i\,\beta,\ 1-x_i\,\beta)$, y no en el intervalo $(-\infty, \infty)$. Y esto hace que los test de significación de variables, del tipo t de Student sean necesariamente sesgados.

Además, el modelo lineal resulta inadecuado para predecir probabilidades, pues nada nos asegura que el valor calculado $\hat{y}_i = x_i\,\hat{\beta}$ (estimación de la probabilidad condicional de realización del suceso) esté comprendido en el intervalo (0,1), que es la característica de una probabilidad.



Esto es un hándicap serio, sobre todo para realizar predicciones, por ello es preferible estimar una función no lineal que tome sus valores en el intervalo (0,1), en la cual y_i sea una función no decreciente de $x_i\,\beta$.

2. LAS TECNICAS DE ESTIMACION DE DATOS AGRUPADOS

El modelo de regresión lineal presenta numerosos inconvenientes, entre los más importantes figura el hecho de no imponer al valor calculado del modelo estar comprendido en el intervalo (0,1). Una solución eventual, para evitar este problema, consiste en representar el fenómeno explicado por una forma no lineal que varíe en el intervalo (0,1) y realizar después una transformación del mismo, de manera que vuelva a presentar la expresión lineal usual. En este principio se basan las técnicas de datos agrupados, las cuales poseen el doble inconveniente de reducir considerablemente la información disponible y de imponer los criterios subjetivos del analista.

2.1. El análisis probit (2)

Este método fue utilizado primeramente por los biólogos, para explicar los efectos de un veneno en una población de animales o de plantas. Las respuestas de los individuos a unos determinados estímulos que examinan estos investigadores, son comparables a ciertas reacciones de los agentes económicos. Por ejemplo, para cada familia, existe un determinado nivel de renta por debajo del cual la familia no poseerá automóvil y por encima del cual sí lo poseerá. Este nivel, que el biólogo llamaría tolerancia, es una variable aleatoria y puede así caracterizarse por una función de densidad de probabilidad. Si un nivel de renta x_0 se atribuye a una familia y si f (x) representa la función de densidad de la tolerancia, entonces la probabilidad de realización del suceso para esta familia vendría dada por:

$$P = \int_0^{x_0} f(x)\,dx$$

En el caso concreto del análisis probit, supongamos una población de animales en la que estudiamos los efectos de diferentes dosis de veneno. Esta población se reparte en G grupos n_i; $i = 1, \ldots, G$. A cada animal de un grupo n_i se le administrará una dosis t_i de veneno. Diremos que $y_{ij} = 1$ si el animal j del grupo i muere como consecuencia de la inyección t_i de veneno y en caso contrario $y_{ij} = 0$. Sea Prob $(y_{ij} = 1)$ la probabilidad de que un animal del grupo i no sobreviva y $(a+bt_i)$ el nivel de veneno a partir del cual el animal muere. Si representamos esta probabilidad por una función no decreciente de t_i, tendremos:

(2) El término "Probit" fue propuesto por C.I. BLISS (1934) y es la contracción de Probability Unit".

$$\text{Prob } (y_{ij} = 1) = F\ (a+bt_i)$$

siendo F una función de distribución de probabilidad, por lo que p_i ha de estar necesariamente comprendida en el intervalo (0,1).

De manera más general, sea $x_i\,\beta$ el nivel a partir del cual el suceso explicado se realiza. El modelo se escribe entonces:

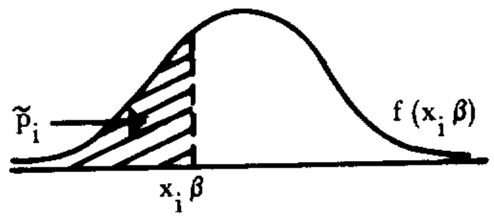
$$\text{Prob } (y_{ij} = 1) = F\ (x_i\,\beta)$$

El análisis Probit consiste en utilizar para F la función de distribución de una variable aleatoria normal. Para ello necesitamos antes la definición de datos agrupados, para así reemplazar la variable dicotómica explicada por su frecuencia de aparición en cada grupo inicialmente definido.

Sea $\tilde{p}_i$ el estimador de la probabilidad de realización del suceso explicado en el interior de cada grupo i, así:

$$\tilde{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \qquad i = 1, \ldots, G$$

Sea $f(x_i\,\beta)$ la función de densidad de probabilidad del nivel de reacción del individuo i; $\tilde{p}_i$ puede entonces representarse de la forma siguiente:



En el caso de una variable aleatoria normal, la función de distribución es:

$$\tilde{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i\beta} e^{(-\frac{1}{2}u^2)}\, du$$

Llamando $z_i = x_i\beta$, queda:

$$\tilde{p}_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{(-\frac{1}{2}u^2)}\, du$$

con lo cual la variable $\tilde{p}_i$ tomará valores comprendidos en el intervalo (0,1).

Es además posible adoptar el modelo de regresión lineal para determinar z_i:

$$z_i = x_i\beta + \epsilon_i \qquad i = 1, \ldots, G$$

con las hipótesis:

$$E(\epsilon_i) = 0, \quad \forall i$$

$$\begin{aligned} E(\epsilon_i \epsilon_j) &= \sigma^2 && \text{para } i = j \\ &= 0 && \text{para } i \neq j \end{aligned}$$

y estimar dicho modelo por el método de los mínimos cuadrados ordinarios. Sin embargo, es necesario que exista, en cada grupo de individuos anteriormente definido, al menos un individuo para el cual el suceso explicado se realice, en caso contrario z_i no sería determinado.

Aparte de la restricción anterior, el análisis Probit tiene dos grandes inconvenientes: el cálculo de la integral anterior no es simple y se necesita acudir a un procedimiento numérico para resolverla. Por otra parte, el análisis de los fenómenos económicos no es análogo al de los fenómenos biológicos y en ningún caso el económetra dispone de datos agrupados; entonces, aparte de lo arbitrario de esta técnica, es extremadamente difícil de agregar los individuos según valores de sus variables y tanto más cuando su número es elevado.

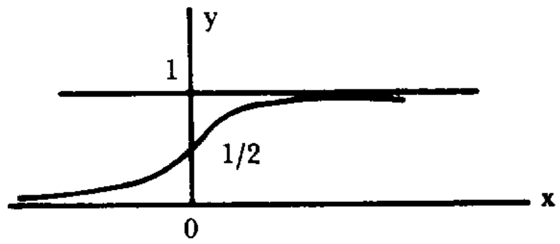
2.2. El análisis Logit (3)

Mientras el análisis Probit utiliza la función de distribución de una variable aleatoria normal, para limitar el valor calculado de la variable dependiente al intervalo (0,1), el análisis Logit utiliza la función de distribución logística estandarizada. Esta función, conocida también con el nombre de ley de Verhulst, se escribe de la siguiente manera:

$$y = \frac{1}{1+e^{-bx}} \qquad -\infty < x < \infty$$

Esta función, empleada tanto en el caso de datos agrupados como en el caso de datos individuales, como veremos en la siguiente sección, merece una atención particular debido a su simplicidad, comparativamente a otras funciones utilizadas en la estimación de modelos de variable dependiente cualitativa.

La representación gráfica de esta función es:



(3) El término "Logit" fue propuesto por J. Berkson (1944) y es la contracción de "Logistic Unit".

El modelo que queremos estimár tiene la forma:

$$\text{Prob}\ (y_{ij} = 1) = F\ (x_i\, \beta)$$

Por lo tanto, si se adopta la función logística como representante del fenómeno estudiado, tendremos entonces:

$$\text{Prob}\ (y_{ij} = 1) = \frac{1}{1+e^{-x_i \beta}}$$

y podremos escribir:

$$1 - \text{Prob}\ (y_{ij} = 1) = 1 - \frac{1}{1+e^{-x_i \beta}} = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}}$$

o bien,

$$\frac{\text{Prob}\ (y_{ij}=1)}{1-\text{Prob}\ (y_{ij} = 1)} = e^{x_i \beta}$$

y tomando logaritmos en ambos miembros tenemos:

$$\text{Ln}\ \frac{\text{Prob}\ (y_{ij} = 1)}{1-\text{Prob}\ (y_{ij} = 1)} = x_i \beta$$

Con lo que nos encontramos de nuevo con el modelo lineal clásico. Si estimamos la probabilidad de realización del suceso explicado por su frecuencia de aparición, en cada grupo de individuos anteriormente definido, tendremos:

$$\text{Ln}\ \frac{\tilde{p}_i}{1-\tilde{p}_i} = x_i \beta$$

con

$$\tilde{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \qquad i = 1, \ldots, G$$

Y será posible utilizar el método de los mínimos cuadrados ordinarios, para encontrar los estimadores de los parámetros desconocidos del modelo.

Este método, atrayente por su simplicidad, necesita la utilización de datos agrupados. Y aunque sea posible particionar los individuos de una muestra, según los valores de las variables que les son asignados, es preferible desarrollar métodos que utilicen datos individuales, para evitar el doble inconveniente de los grupos vacíos y el de disminuir la información disponible cuando el tamaño de la muestra es reducido.

A estos problemas se añade el de la estimación del modelo, ya que el método se basa en la transformación de un modelo no lineal en uno lineal y esta transformación no sólo afecta a las variables endógena y exógenas, sino también el término de error aleatorio, lo que hace que éste ya no satisfaga las hipótesis habituales. En efecto, tenemos:

$$\tilde{p}_i = \frac{1}{1+e^{-x_i \beta}} + \epsilon_i$$

o bien, con la transformación anterior:

$$\frac{\tilde{p}_i}{1-\tilde{p}_i} = \frac{1+\epsilon_i\,(1+e^{-x_i \beta})}{e^{-x_i \beta} - \epsilon_i\,(1+e^{-x_i \beta})}$$

Y tomando logaritmos:

$$\mathrm{Ln}\ \frac{\tilde{p}_i}{1-\tilde{p}_i} = \mathrm{Ln}\,[1+\epsilon_i\,(1+e^{-x_i\beta})] -$$

$$- \mathrm{Ln}\,[e^{-x_i\beta} - \epsilon_i\,(1+e^{-x_i\beta})]$$

que vemos no es igual a

$$\mathrm{Ln}\ \frac{\tilde{p}_i}{1-\tilde{p}_i} = x_i\beta + \epsilon_i$$

En buena lógica, este modelo no constituye en ningún caso el modelo inicial transformado. Por lo tanto, su estimación no tiene en cuenta las características de la perturbación aleatoria, que se supone representa los factores omitidos en la explicación de la frecuencia de aparición del suceso considerado, con lo cual la estimación de la probabilidad será necesariamente sesgada. Para paliar este inconveniente sería preferible no transformar el modelo, pues así se resolvería el doble problema de la agregación de los individuos y el de los tests para la contrastación de hipótesis.

2.3. Transformaciones de la función de probabilidad lineal

La diferencia esencial entre estos métodos y los precedentes reside en la introducción ex-post de la no linealidad y en la naturaleza de los datos tratados, ya que en ellos no se necesita la definición previa de grupos de individuos, según los valores de las variables exógenas que le sean asignados, sino que se definen los grupos de individuos según el valor calculado de la probabilidad de realización del suceso explicado o según el carácter dicotómico de la variable endógena estudiada.

a) El llamado método de transformación logística de la función de probabilidad lineal, consiste en ordenar en clases los valores de predicción obtenidos de la estimación de la forma lineal:

$$y_i = x_i\beta + \epsilon_i \qquad \text{(II)}$$

y calcular para cada una de ellas la expresión:

$$\hat{L} = \mathrm{Ln}\,\left(\frac{\bar{\hat{y}}}{1-\bar{\hat{y}}}\right)$$

donde $\bar{\hat{y}}$ representa la media de las predicciones de cada intervalo, obtenidas por el método de los mínimos cuadrados generalizados aplicado al modelo anterior (II).

La regresión lineal de la expresión (II) con las marcas de clase, permite relacionar $\hat{L}$ con $\hat{y}$ e indirectamente con el vector de las variables exógenas. De esta manera se expresa entonces la probabilidad del suceso considerado:

$$\widehat{\mathrm{Prob}\,(y_i = 1)} = \frac{1}{1+e^{-\hat{L}_i}}$$

Aunque este método permite no obtener valores de predicción negativos o superiores a uno, sin embargo tampoco es totalmente satisfactorio, pues se basa en la estimación de la función de probabilidad lineal por el método de los mínimos cuadrados generalizados, cuyos tests de significación son necesariamente sesgados como ya vimos, necesitando además la definición arbitraria de intervalos para los valores $\hat{y}_i$.

b) El llamado método de transformación de Warner de la función de probabilidad lineal, consiste en estimar la probabilidad de realización del suceso explicado mediante:

$$\widehat{\mathrm{Prob}\,(y_i = 1)} = \frac{e^{\hat{D}(x_i)}}{1+e^{\hat{D}(x_i)}}$$

donde $\hat{D}$ es el estimador de la función discriminante, expresada en función de $\hat{y}$ de la siguiente forma:

$$\hat{D}(x_i) = K(\hat{y}_i - \hat{\beta}_0)$$

con

$$K = \frac{N_1 + N_2 - 2}{\dfrac{N_1 N_2}{N_1+N_2} - \dfrac{N_1 N_2}{N_1+N_2}(\bar{x}_1 + \bar{x}_2)' \hat{\beta}_\star}$$

y $\hat{\beta}_0$ es el estimador por mínimos cuadrados generalizados de la constante de la función de probabilidad lineal. N_1 es el número de individuos que tienen 1 como valor de la variable endógena, N_2 el número de individuos que tienen 0, $\bar{x}_1$ el vector de las medias de las variables exógenas del grupo de individuos con 1 como valor de la variable endógena, $\bar{x}_2$ lo mismo para el grupo con 0 y $\hat{\beta}_\star$ el vector de los estimadores de las pendientes desconocidas de la función de probabilidad lineal, por el método de los mínimos cuadrados generalizados.

Las mismas críticas que para el método anterior se dirigen al de la transformación de Warner; aunque éste presenta una ligera ventaja respecto al otro, pues no necesita la creación arbitraria de intervalos para los valores $\hat{y}_i$. A pesar de ello, también introduce artificialmente la no linealidad, sin considerar sus efectos sobre los estimadores de los parámetros del modelo, ni sobre la significación de las variables que se suponen determinan el fenómeno explicado.

3. EL MODELO LOGISTICO DE VARIABLE ENDOGENA BINARIA

Para evitar los inconvenientes de los métodos anteriores, es mejor postular, desde el principio, una función no decreciente del fenómeno explicado y estimar los parámetros desconocidos de la misma por los métodos clásicos de inferencia estadística, sin transformación del modelo ni utilización de datos agrupados artificialmente. Sería posible realizar numerosas formulaciones, pero por su simplicidad se impone particularmente una: el modelo logístico. La estimación de tal modelo por el método de la máxima verosimilitud, permite obtener estimadores asintóticamente eficientes y construir tests de contrastación "exactos", para juzgar la significación de las variables que se suponen determinan el fenómeno estudiado.

3.1. Construcción de la función de verosimilitud

El principio del método de máxima verosimilitud, utilizado para la estimación del modelo logístico, es el siguiente: La función de densidad de probabilidad conjunta de la variable aleatoria y, considerada como función de los parámetros desconocidos $\beta' = (\beta_0, \ldots, \beta_k)$, es llamada función de verosimilitud. Sea $L(y, \beta')$ esta función. Cuando se dispone de observaciones de y, Fisher propone estimar el valor β_k desconocido por el valor $\hat{\beta}_k(y)$ del parámetro que maximiza la verosimilitud de la muestra, es decir:

$$L[y, \hat{\beta}_k(y)] \geqslant L[y, \beta_k], \ \forall \beta_k$$

El problema que debemos resolver es el de la especificación de la función de densidad conjunta de las observaciones y_i.

Nuestro modelo a estimar es:

$$y_i = \frac{1}{1+e^{-x_i \beta}} + \epsilon_i$$

con las hipótesis habituales acerca del término de perturbación, es decir:

$$E(\epsilon_i) = 0, \ \forall i$$

$$E(\epsilon_i \epsilon_j) = y_i (1-y_i) \qquad \text{para } i = j$$
$$= 0 \qquad \text{para } i \neq j$$

Haciendo uso de la hipótesis de que la esperanza matemática de la perturbación aleatoria es nula, la probabilidad de realización del suceso explicado es igual a:

$$\text{Prob}\ (y_i = 1) = E(y_i \mid x_i \beta) = \frac{1}{1+ e^{-x_i \beta}}$$

Por otra parte, la perturbación no puede tomar más que dos valores:

$$\epsilon_i = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}} \qquad \text{si } y_i = 1$$

$$\epsilon_i = -\frac{1}{1+e^{-x_i \beta}} \qquad \text{si } y_i = 0$$

Sabiendo que:

$$E(\epsilon_i) = \text{Prob}\ (\epsilon_i = \frac{e^{-x_i \beta}}{1+ e^{-x_i \beta}})\ (\frac{e^{-x_i \beta}}{1+e^{-x_i \beta}}) +$$

$$+ \text{Prob}\ (\epsilon_i = -\frac{1}{1+e^{-x_i \beta}})\ (-\frac{1}{1+e^{-x_i \beta}}) = 0$$

y que

$$\text{Prob}\ (\epsilon_i = \frac{e^{-x_i \beta}}{1 + e^{-x_i \beta}}) + \text{Prob}\ (\epsilon_i = -\frac{1}{1+e^{-x_i \beta}}) = 1$$

Por un cálculo análogo al efectuado para la función de probabilidad lineal, obtenemos las siguientes expresiones para las probabilidades de realización de la perturbación ϵ_i:

$$\text{Prob}\ (\epsilon_i = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}}) = \frac{1}{1+e^{-x_i \beta}}$$

y

$$\text{Prob}\ (\epsilon_i = -\frac{1}{1+e^{-x_i \beta}}) = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}}$$

De lo que resulta:

$$f\ (\epsilon_i) = \frac{1}{1+ e^{-x_i \beta}} \qquad \text{cuando } y_i = 1$$

$$f\ (\epsilon_i) = \frac{e^{-x_i \beta}}{1+ e^{-x_i \beta}} \qquad \text{cuando } y_i = 0$$

Con lo cual deducimos que:

$$f\ (y_i) = \frac{1}{1+e^{-x_i \beta}} \qquad \text{para } y_i = 1$$

y

$$f\ (y_i) = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}} \qquad \text{para } y_i = 0$$

Cada variable aleatoria y_i es una variable aleatoria binomial, cuya función de densidad de probabilidad se escribirá así:

$$f(y_i) = \left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i}\left(\frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right)^{1-y_i}$$

Como y_i es independiente de y_j, la función de densidad de probabilidad conjunta no será más que el producto de las funciones de densidad de probabilidad individuales. Así, la función de verosimilitud de la muestra tendrá la siguiente expresión:

$$\text{(III)}\ L(\beta_0, \ldots, \beta_K \mid y_1 \ldots y_i \ldots y_n) =$$

$$= \prod_{i=1}^{n}\left(\frac{1}{1+e^{-x_i\beta}}\right)^{y_i}\left(\frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}\right)^{1-y_i}$$

3.2. Estimación y tests de hipótesis

La estimación del modelo por el método de máxima verosimilitud, consiste en maximizar la función $L(\beta_0, \ldots, \beta_k \mid y_1 \ldots \ldots y_i \ldots y_n)$ respecto a todos los parámetros desconocidos β_k. La condición para tener un máximo es que las derivadas primeras de la función de verosimilitud, respecto a los parámetros desconocidos, sean nulas.

Habitualmente, la resolución del sistema lineal de ecuaciones normales permite deducir estos estimadores, en el caso lineal. Es evidente que en este caso, las ecuaciones no son lineales en los parámetros, por lo que la resolución del sistema de ecuaciones

normales no es tan simple. Sólamente un método de optimización numérica permitirá obtener los estimadores de los parámetros desconocidos.

Como la función que queremos maximizar $L(\beta_0, \ldots, \beta_k \mid y_1 \ldots y_i \ldots y_n)$ es convexa, estamos seguros de encontrar un máximo global, de tal modo que los estimadores de los parámetros desconocidos del modelo posean todas las características de los estimadores máximo-verosímiles. Así, estos estimadores son:

a) Consistentes, es decir que:

$$p \lim \hat{\beta}_k = \beta_k$$

b) Asintóticamente eficientes. Cualquier otro estimador consistente $\beta^{\#}_k$ de β_k posee una variancia asintótica superior a la de $\hat{\beta}_k$ (4).

c) Además, estos estimadores son asintóticamente normales.

Una vez encontrada, mediante un método de optimización, la estimación de los parámetros desconocidos del modelo, podremos juzgar el poder explicativo de una variable cualquiera x_k, procediendo de dos formas diferentes. La primera consiste en calcular el cociente entre el estimador y su variancia asintótica, definida por:

$$\text{Var. asintótica}\ (\hat{\beta}_k) = -\left(\frac{\delta^2 L}{\delta \beta_k^2} \Big/ \beta_k = \hat{\beta}_k\right)^{-1}$$

y comparar ese cociente a una t de Student.

(4) La variancia asintótica de los estimadores máximo-verosímiles se define como:

$$\text{Var. Asintótica}\ (\hat{\beta}_k) = \frac{1}{n} \lim_{n \to \infty} n\ \text{Var}\ (\hat{\beta}_k)$$

La segunda consiste en utilizar el test de la razón de verosimilitudes, esto es:

$$\lambda = \frac{L_e}{L}$$

donde L representa el valor de la función de verosimilitud en el punto $\hat{\beta}$ y L_e representa el valor de la función de verosimilitud en el punto $\hat{\beta}'_e = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{k-1}, 0)$

como

$$-2 \text{ in } \lambda \underset{\text{asint.}}{\sim} \chi 2 \quad (1)$$

La comparación de la cantidad $-2 \ln \lambda$ con una χ^2 teórica, permite entonces rechazar o aceptar la hipótesis en cuestión.

4. EL MODELO LOGISTICO DE VARIABLE DEPENDIENTE MULTIPLE

Hasta ahora hemos estudiado el modelo de regresión cuya variable dependiente es binaria. En este caso se revela el modelo logístico, estimado por el método de la máxima verosimilitud, como el que procura los mejores resultados, tanto desde el punto de vista de la calidad de los estimadores de los parámetros desconocidos, como desde el punto de vista de las predicciones obtenidas.

Naturalmente hay muchas situaciones en las cuales la variable dependiente cualitativa de un modelo es múltiple, es decir, que admite más de dos modalidades. Nos proponemos en esta sección presentar la generalización del modelo logístico de la variable dependiente dicotómica al modelo de la variable múltiple.

4.1. Presentación del modelo

Imaginemos un bachiller i decidido a continuar estudios su-

periores universitarios. Este individuo deberá hacer una elección para decidir el lugar donde realizará sus estudios. Entre un cierto número de Universidades deberá elegir una. Si solamente existieran dos Universidades, U_1 y U_2, estaríamos ante el caso de la variable binaria ya estudiada:

$y_i = 1$ si el individuo elige la Universidad U_1

$y_i = 0$ si el individuo elige la Universidad U_2

Imaginemos que la variable y_i está determinada por k variables exógenas independientes y fijas x (binarias o no). Simbolizaremos por y_i la elección del individuo i, i = 1, . . . , n. La probabilidad de dirigirse a la Universidad U_1 será:

$$\text{Prob } (y_i = U_1) = \frac{1}{1+e^{-x_i \beta}}, \quad i = 1, \ldots, n \quad \text{(III)}$$

donde β es el vector de orden (k+1, 1) de parámetros desconocidos.

También hubiese sido posible estimar el modelo:

$$\text{Prob } (y_i = U_2) = \frac{1}{1+e^{-x_1 \beta^{\#}}}, \quad i = 1, \ldots, n$$

es decir, el suceso contrario, donde $\beta^{\#}$ es el vector de orden (k+1, 1) compuesto por: $\beta^{\#\prime} = (\beta^{\#}_0, \beta^{\#}_1, \ldots, \beta^{\#}_k)$. En este caso:

$y_i = 1$ si el individuo i elige la Universidad U_2

$y_i = 0$ si el individuo i elige la Universidad U_1

Naturalmente en este segundo modelo, los valores absolutos de los parámetros son los mismos que los de los parámetros del modelo inicial, pero sus signos están invertidos; verificándose:

$$\beta_k + \beta^{\#}_k = 0 \qquad \forall k$$

puesto que

$$\text{Prob}\,(y_i = U_2) = 1 - \text{Prob}\,(y_i = U_1)$$

Es posible escribir el modelo (III) de la forma siguiente:

$$\text{Prob}\,(y_i = U_1) = \frac{e^{x_i \beta}}{e^{x_i \beta} + e^{x_i \beta^{\#}}}$$

y puesto que $\beta = -\beta^{\#}$, tendremos:

$$\text{Prob}\,(y_i = U_1) = \frac{1}{1+e^{-2x_i \beta}}$$

La función de distribución logística univariante se escribe:

$$f\,(t_1) = \frac{1}{1+e^{-t_1}} \qquad -\infty < t_1 < \infty$$

entonces, las probabilidades Prob $(y_i = U_1)$ y Prob $(y_i = U_2)$ se obtienen sustituyendo $t_1 = 2x_i \beta$.

Supongamos ahora el caso más realista en que el individuo i debe elegir entre un conjunto de Q Universidades: $U_1, U_2, \ldots, U_q$. Aquí, el suceso

$$E = \left\{ \text{lugar de realización de los estudios superiores} \right\}$$

es una variable múltiple que comporta tantas modalidades como Universidades existan en el territorio nacional.

Sea entonces

$$P_{ij} = \text{Prob}\,(y_i = U_j)$$

la probabilidad de que el individuo i elija la Universidad j. En este caso tendremos:

$$\sum_{j=1}^{Q} P_{ij} = 1 \qquad \forall i,\ i = 1, \ldots, n$$

Utilizaremos la función logística estandarizada, para expresar la relación entre la elección del individuo i y el vector x_i de las variables exógenas que se suponen determinan su preferencia. En el caso multivariante, la función de distribución logística se escribe de la manera siguiente:

$$f(t_1, \ldots, t_m) = \frac{1}{1 + \sum_{j=1}^{m} e^{-t_j}}, \quad -\infty < t_j < \infty$$

En nuestro caso, por analogía con el modelo binario, es posible escribir la probabilidad de la forma siguiente:

$$P_{ij} = \frac{e^{x_i \beta_j}}{\sum_{k=1}^{Q} e^{x_i \beta_k}} \qquad j = 1, \ldots, Q;\ i = 1, \ldots, n$$

con

$$\sum_{j=1}^{Q} P_{ij} = 1$$

y

$$\sum_{j=1}^{Q} \beta_j = 0$$

donde:

x_i es el vector de orden $(1, k+1)$ de las variables explicativas asociadas al individuo i

β_j es el vector de orden $(k+1, 1)$ de los parámetros desconocidos asociados a la modalidad j del suceso considerado.

4.2. Estimación del modelo

A semejanza del modelo de variable dependiente binaria, el modelo de variable endógena múltiple se estima también por el método de máxima verosimilitud. Mientras que en el primer caso debimos estimar solamente una función de probabilidad y un sólo vector de parámetros desconocidos, sin embargo en el segundo será necesario estimar tantas funciones de probabilidad como modalidades pueda tomar la variable cualitativa múltiple y en consecuencia habrá que encontrar (Q x K) parámetros desconocidos.

En el caso binario presentado precedentemente teníamos:

$$\text{Prob}\ (y_i = U_1) = \frac{1}{1+e^{-x_i \beta}}$$

y

$$\text{Prob}\ (y_i = U_2) = \frac{e^{-x_i \beta}}{1+e^{-x_i \beta}}$$

La función de verosimilitud se escribe en este caso como en (III).

Llamando P_{i1} a la probabilidad de que el individuo i se dirija a la Universidad U_1 y P_{i2} a la probabilidad de que se dirija a la Universidad 2, tendremos entonces:

$$L(\beta_0, \ldots, \beta_k \mid y_1, \ldots, y_n) = \prod_{i=1}^{n} P_{i1}^{y_i} P_{i2}^{1-y_i}$$

Si el individuo i se dirige a la Universidad U_1 escribiremos:

$$v_{i1} = 1 \quad y \quad v_{i2} = 0$$

Si por el contrario el individuo i se dirige a la Universidad U_2, escribiremos:

$$v_{i1} = 0 \quad y \quad v_{i2} = 1$$

La función de verosimilitud del modelo podremos entonces expresarla así:

$$L(\beta_0, \ldots\ldots, \beta_k \mid y_1 \ldots y_n) = \prod_{i=1}^{n} P_{i1}^{v_{i1}} P_{i2}^{v_{i2}}$$

o bien, de forma más condensada:

$$L(\beta_0, \ldots\ldots, \beta_k \mid y_1 \ldots y_n) = \prod_{i=1}^{n} \prod_{j=1}^{2} P_{ij}^{v_{ij}}$$

Para el caso de Q modalidades tendremos:

$$L(\beta_0, \ldots\ldots, \beta_k \mid y_1 \ldots y_n) = \prod_{i=1}^{n} \prod_{j=1}^{Q} P_{ii}^{v_{ij}}$$

La estimación del modelo por el método de la máxima verosimilitud, consiste en maximizar la función $L(\beta_0 \ldots \beta_k \mid y_1 \ldots$

$\ldots y_n)$ con relación a todos los parámetros desconocidos del modelo. Sin embargo, para que todos los (Q x K) parámetros esten definidos, es necesario maximizar esta función bajo la restricción:

$$\sum_{j=1}^{Q} \beta_j = 0$$

donde β_j es el vector de parámetros de las variables exógenas asociadas a la j-ésima modalidad del fenómeno explicado.

La condición para tener un máximo es que las derivadas primeras de la función de verosimilitud respecto a los parámetros sean nulas. Como en el caso binario, al no ser lineales las ecuaciones normales, es necesario recurrir a un proceso de optimización. Sin embargo, puesto que la función es convexa, estamos seguros de encontrar un máximo global y que entonces los parámetros desconocidos del modelo posean todas las características de los estimadores máximo-verosímiles.

Los tests de hipótesis se construyen de igual forma que en el caso binario, esto es, o bien comparando la razón de verosimilitudes a una χ^2 teórica con un grado de libertad, o bien comparando el cociente entre el estimador del parámetro y su variancia asintótica a una t de Student.

En un modelo lineal, el coeficiente de correlación múltiple proporciona una medida de la mejor o peor relación (lineal) que existe entre las variables explicativas en su conjunto y la variable explicada. Ahora bien, en un modelo no lineal como el modelo logístico, no podemos calcular tal coeficiente, por eso es necesario efectuar un test de significación del conjunto de variables exógenas, que nos permita juzgar si las variables consideradas determinan de manera significativa la probabilidad de realización del suceso explicado.

5. EL MODELO LOGISTICO DE ECUACIONES SIMULTANEAS DE VARIABLES ENDOGENAS CUALITATIVAS

La introducción de variables endógenas cualitativas en un modelo de ecuaciones simultáneas, plantea los mismos problemas de estimación que los encontrados en un modelo simple de variable explicada cualitativa. Aquí también es necesario reemplazar la forma lineal del modelo por una no lineal, tal como la forma logística, cuyos parámetros serán estimados por el método de máxima verosimilitud con información completa.

Para facilitar la exposición, consideraremos un modelo que contenga solamente dos variables endógenas cualitativas. Primero efectuaremos el examen detallado del modelo simultáneo de variables dependientes binarias y después generalizaremos al modelo de variables dependientes múltiples.

5.1. Presentación del modelo

Un modelo econométrico es una formalización de un modelo económico que expresa, mediante un conjunto de ecuaciones, ciertas variables económicas en función de ellas mismas y de otras variables económicas o no. La característica de un sistema simultáneo reside en el hecho de que una variable explicada en una ecuación aparece como variable explicativa en otra ecuación.

Tengamos el suceso:

$$E = \{\text{empleo ocupado por el individuo i}\}$$

Consideraremos solamente dos tipos de empleo: los técnicos y los administrativos, en cuyo caso podremos definir una variable dicotómica de la manera siguiente:

$y_{i1} = 1$ si el individuo i ocupa un empleo técnico

$y_{i1} = 0$ si el individuo i ocupa un empleo administrativo.

Supongamos que esta variable esté determinada por k variables exógenas x, independientes y fijas, y por una variable binaria y_2 que simboliza el suceso:

$$E' = \left\{ \text{sector de actividad del individuo i} \right\}$$

que codificamos así:

$y_{i2} = 1$ si el sector de actividad del individuo i es el sector industrial

$y_{i2} = 0$ si es el sector servicios.

Tendremos entonces el modelo siguiente:

$$y_{i1} = f\,(x_i, y_{i2}, \epsilon_{i1})$$

y bajo la hipótesis de una representación logística del fenómeno, escribiremos:

$$y_{i1} = \frac{1}{1+e^{-x_i\beta-\alpha y_{i2}}} + \epsilon_{i1} \qquad i = 1, \ldots n$$

donde

x_i es el vector de orden (1, k+1) de variables exógenas que determinan la variable binaria y_{i1}.

β es el vector de orden (k+1, 1) de los parámetros desconocidos de esas variables.

y_{i2} es la variable dicotómica que representa el sector de actividad del individuo i.

α el coeficiente de esa variable.

ϵ_{i1} un término de perturbación aleatoria.

Bajo la hipótesis

$$E\,(\epsilon_{i1}) = 0 \qquad \forall i$$

tendremos

$$E\,(y_{i1} \mid x_i, y_{i2}) = \text{Prob}\,(y_{i1} = 1 \mid x_i, y_{i2}) = \frac{1}{1+e^{-x_i\beta-\alpha y_{12}}}$$

Imaginemos que queremos explicar la variable binaria y_2; esta variable viene determinada por z variables exógenas, además de por la variable endógena y_1 (puesto que la probabilidad de trabajar en un sector de actividad determinado está parcialmente condicionada por la probabilidad de ejercer un empleo u otro). Si seguimos considerando la hipótesis de una relación logística para representar el fenómeno, tendremos:

$$y_{i2} = \frac{1}{1+e^{-z_i\gamma - s\,y_{i1}}} + \epsilon_{i2}$$

donde

z_i es el vector de orden (1, d+1) de variables exógenas que determinan la variable binaria y_{i2}

γ el vector de orden (d+1, 1) de parámetros asociados a esas variables.

y_{i1} la variable binaria representante del empleo ocupado por el individuo i

s el coeficiente que mide el efecto de esta variable sobre el sector de actividad

ϵ_{i2} un término de error aleatorio

Bajo la hipótesis $E(\epsilon_{i2}) = 0 \quad \forall i$
y considerando momentáneamente y_{i1} como una variable exógena, tenemos:

$$E(y_{i2} \mid z_i, y_{i1}) = \text{Prob}(y_{i2} = 1 \mid z_i, y_{i1}) = \frac{1}{1+e^{-z_i\gamma - sy_{i1}}}$$

Por la simultaneidad del problema, es imposible estimar los parámetros desconocidos a partir de cada ecuación considerada independientemente de la otra. Las estimaciones deben efectuarse a partir del conjunto de ecuaciones, es decir, que deberemos considerar el modelo siguiente:

$$\text{Prob}(y_{i1} = 1 \mid y_{i2}) = \frac{1}{1+e^{-x_i\beta - \alpha y_{i2}}}$$

$$\text{Prob}(y_{i2} = 1 \mid y_{i1}) = \frac{1}{1+e^{-z_i\gamma - sy_{i1}}}$$

Este sistema de variables endógenas interdependientes no es un modelo general de ecuaciones simultáneas, pero nos hemos servido de él con el fin de simplificación, ya que un sistema que contenga más de dos variables endógenas resulta muy complejo.

5.2. Estimación del modelo

Dado el carácter interdependiente de las variables endógenas, no es necesario estimar (k+d+2) parámetros sino sólo (k+d+1), pues los coeficientes de las variables endógenas en cada ecuación son idénticos. En efecto, las probabilidades condicionales de los diferentes sucesos son iguales a:

$$\text{Prob}(y_{i1} = 1 \mid y_{i2} = 0) = \frac{\text{Prob}(y_{i1} = 1, y_{i2} = 0)}{\text{Prob}(y_{i2}=0)} = $$

$$= \frac{1}{1+e^{-x_i\beta}}$$

$$\text{Prob}(y_{i1} = 0 \mid y_{i2} = 0) = \frac{\text{Prob}(y_{i1} = 0, y_{i2} = 0)}{\text{Prob}(y_{i2} = 0)} = $$

$$= \frac{e^{-x_i\beta}}{1+e^{-x_i\beta}}$$

$$\text{Prob}(y_{i1} = 1 \mid y_{i2} = 1) = \frac{\text{Prob}(y_{i1} = 1, y_{i2} = 1)}{\text{Prob}(y_{i2} = 1)} = $$

$$= \frac{1}{1+e^{-x_i\beta-\alpha}}$$

$$\text{Prob}(y_{i1} = 0 \mid y_{i2} = 1) = \frac{\text{Prob}(y_{i1} = 0, y_{i2} = 1)}{\text{Prob}(y_{i2} = 1)} = $$

$$= \frac{e^{-x_i\beta-\alpha}}{1+e^{-x_i\beta-\alpha}}$$

De donde obtenemos:

$$e^{-x_i\beta} = \frac{\text{Prob}(y_{i1} = 0, y_{i2} = 0)}{\text{Prob}(y_{i1} = 1, y_{i2} = 0)}$$

$$e^{-x_i\beta-\alpha} = \frac{\text{Prob }(y_{i1} = 0, y_{i2} = 1)}{\text{Prob }(y_{i1} = 1, y_{i2} = 1)}$$

y reemplazando en esta última expresión $e^{-x_i\beta}$ por la relación de las probabilidades conjuntas correspondientes, obtenemos:

$$e^{\alpha} = \frac{\text{Prob }(y_{i1} = 0, y_{i2} = 0)}{\text{Prob }(y_{i1} = 1, y_{i2} = 0)} : \frac{\text{Prob }(y_{i1} = 0, y_{i2} = 1)}{\text{Prob }(y_{i1} = 1, y_{i2} = 1)}$$

Es decir:

$$e^{\alpha} = \frac{\text{Prob }(y_{i1} = 1, y_{i2} = 1)\ \text{Prob }(y_{i1} = 0, y_{i2} = 0)}{\text{Prob }(y_{i1} = 0, y_{i2} = 1)\ \text{Prob }(y_{i1} = 1, y_{i2} = 0)}$$

De igual manera construiremos las probabilidades condicionales para la variable y_{i2} y por un cálculo análogo al anterior obtenemos:

$$e^{s} = \frac{\text{Prob }(y_{i2} = 1, y_{i1} = 1)\ \text{Prob }(y_{i2} = 0, y_{i1} = 0)}{\text{Prob }(y_{i2} = 0, y_{i1} = 1)\ \text{Prob }(y_{i2} = 1, y_{i1} = 0)}$$

De donde:

$$\alpha = s$$

Por lo cual el sistema de ecuaciones simultáneas puede escribirse:

$$\text{Prob }(y_{i1} = 1 \mid y_{i2}) = \frac{1}{1+e^{-x_i\beta-\alpha y_{i2}}}$$

$$\text{Prob }(y_{i2} = 1 \mid y_{i1}) = \frac{1}{1+e^{-z_i\gamma-\alpha y_{i1}}}$$

El modelo podrá estimarse por el método de máxima verosimilitud con información completa, para ello antes se necesita construir la función de verosimilitud del sistema. Esta función no es más que el producto de las probabilidades conjuntas individuales de realización de los sucesos explicados.

Así, la probabilidad del suceso $y_{i1} = 0$ e $y_{i2} = 0$ es igual a:

$$\text{Prob }(y_{i1} = 0, y_{i2} = 0) = \text{Prob }(y_{i1} = 0 \mid y_{i2} = 0)\,\text{Prob }(y_{i2} = 0)$$

Como

$$\text{Prob }(y_{i2} = 0) = \text{Prob }(y_{i2}=0 \mid y_{i1}=0) + \text{Prob }(y_{i2}=0 \mid y_{i1}=1) =$$

$$= \frac{1}{1+e^{z_i\gamma}} + \frac{1}{1+e^{z_i\gamma+\alpha}}$$

y

$$\text{Prob }(y_{i1} = 0 \mid y_{i2} = 0) = \frac{1}{1+e^{x_i\beta}}$$

Tenemos entonces:

$$\text{Prob }(y_{i1} = 0, y_{i2} = 0) = \frac{1}{1+e^{x_i\beta}} \left(\frac{1}{1+e^{z_i\gamma}} + \frac{1}{1+e^{z_i\gamma+\alpha}} \right)$$

Y de la misma manera obtendremos:

$$\text{Prob}\ (y_{i1} = 1, y_{i2} = 0) = \frac{e^{x_i\beta}}{1+e^{x_i\beta}} \left(\frac{1}{1+e^{z_i\gamma}} + \frac{1}{1+e^{z_i\gamma+\alpha}} \right)$$

$$\text{Prob}\ (y_{i1} = 0, y_{i2} = 1) = \frac{1}{1+e^{x_i\beta+\alpha}} \left(\frac{e^{z_i\gamma}}{1+e^{z_i\gamma}} + \frac{e^{z_i\gamma+\alpha}}{1+e^{z_i\gamma+\alpha}} \right)$$

$$\text{Prob}\ (y_{i1} = 1, y_{i2} = 1) = \frac{e^{x_i\beta+\alpha}}{1+e^{x_i\beta+\alpha}} \left(\frac{e^{z_i\gamma}}{1+e^{z_i\gamma}} + \frac{e^{z_i\gamma+\alpha}}{1+e^{z_i\gamma+\alpha}} \right)$$

Es decir, si llamamos i al índice del individuo y p, q a los índices de los eventos, el conjunto de los individuos que tengan el evento p, q será:

$$\vartheta_{p,q} = (i \mid y_{i1} = p, y_{i2} = q) \qquad p, q = 0,1; \quad i = 1, \ldots, n$$

La función de verosimilitud del sistema es entonces igual a:

$$L(\beta, \gamma, \alpha) = \prod_{p=0}^{1} \prod_{q=0}^{1} \prod_{i \in \vartheta_{p,q}} \text{Prob}\ (y_{i1} = p, y_{i2} = q)$$

La estimación de los parámetros desconocidos del sistema por el método de la máxima verosimilitud, consiste en maximizar la función L (β, γ, α) respecto a todos los parámetros desconocidos del modelo. Para ello hay que recurrir a un proceso de optimización numérica como en los casos anteriormente estudiados. Para obtener los tests para la contrastación de hipótesis lo más directo es utilizar la razón de verosimilitudes, ya definida anteriormente.

5.3. Generalización del caso múltiple

Si en vez de considerar dos categorías de empleo, en el ejemplo anterior, tenemos p y en vez de dos sectores de actividad suponemos existen q, las variables cualitativas y_{i1} e y_{i2} son variables múltiples con p y q modalidades respectivamente. En estas condiciones el sistema de ecuaciones simultáneas se escribe:

$$\text{Prob}\ (y_{i1} = r \mid y_{i2}) = \frac{1}{1+e^{-x_i\beta_r - \sum_{s=1}^{q} \alpha_{sr}\, y_{i2}{}^s}} \qquad r=1, \ldots p$$

$$\text{Prob}\ (y_{i2} = s \mid y_{i1}) = \frac{1}{1+e^{-z_i\gamma_s - \sum_{r=1}^{p} \alpha_{rs}\, y_{i1}{}^r}} \qquad s=1, \ldots q$$

donde:

x_i y z_i representan respectivamente los vectores de orden

(1, k+1) y (1, d+1) de las variables exógenas asociadas al individuo i

β_r es el vector de orden (k+1, 1) de parámetros asociados a la modalidad r-ésima de la variable y_{i1}

α_{sr} la influencia de la modalidad s-ésima de la variable y_{i2} en la modalidad r-ésima de la variable y_{i1}

γ_s el vector de orden (d+1, 1) de parámetros asociados a la modalidad s-ésima de la variable y_{i2}

y α_{rs} la influencia de la modalidad r-ésima de la variable y_{i1} en la modalidad s-ésima de la variable y_{i2}

Luego hay que estimar (kxp) parámetros β, (1xq) parámetros γ y (pxq) parámetros α, dado el carácter interdependiente de las variables y_{i1} e y_{i2}.

Como antes, habrá que encontrar los estimadores de los parámetros desconocidos del sistema y lo haremos mediante el método de máxima verosimilitud con información completa, habiendo definido antes, para ello, las funciones de probabilidades individuales conjuntas de realización de los diferentes sucesos.

Según el teorema de las probabilidades condicionales, la probabilidad de realización del suceso $(y_{i1} = r, y_{i2} = s)$ es igual a:

$$\text{Prob}\,(y_{i1} = r, y_{i2} = s) = \text{Prob}\,(y_{i1} = r \mid y_{i2} = s)\ \text{Prob}\,(y_{i2} = s)$$

Como

$$\text{Prob}\,(y_{i1} = r \mid y_{i2} = s) = \frac{1}{1+e^{-x_i\beta_r-\alpha_{sr}}}$$

y

$$\text{Prob}\,(y_{i2} = s) = \sum_{r=1}^{p} \text{Prob}\,(y_{i2} = s \mid y_{i1} = r)$$

siendo además:

$$\text{Prob}\,(y_{i2} = s \mid y_{i1} = r) = \frac{1}{1+e^{-z_i\gamma_s-\alpha_{sr}}}$$

Entonces resulta de todo ello:

$$\text{Prob}\,(y_{i1} = r, y_{i2} = s) = \frac{e^{x_i\beta+\alpha_{sr}}}{1+e^{x_i\beta_r+\alpha_{sr}}}$$

$$\sum_{r=1}^{p} \frac{e^{z_i\gamma_s+\alpha_{sr}}}{1+e^{z_i\gamma_s+\alpha_{sr}}}$$

Las probabilidades de realización de los demás sucesos conjuntos se obtiene de igual forma. No sería cuestión de enumerar todas esas probabilidades, así que construiremos la función de verosimilitud del sistema a partir de la probabilidad conjunta del suceso general:

$$\vartheta_{rs} = (i \mid y_{i1} = r, y_{i2} = s) \qquad r = 1, \ldots, p; \quad s = 1, \ldots, q$$

La función de verosimilitud del sistema es entonces igual a:

$$L\,(\beta, \gamma, \alpha) = \prod_{r=1}^{p} \prod_{s=1}^{q} \prod_{i\in\vartheta_{rs}} \text{Prob}\,(y_{i1} = r, y_{i2} = s)$$

Por un proceso de optimización numérica será posible encontrar los estimadores de los parámetros desconocidos del sistema, además de construir los tests para la contrastación de hipótesis.

6. CONCLUSION

Elegir la habitual forma lineal, para representar la relación entre una variable dependiente cualitativa y un cierto número de variables exógenas, hemos visto que produce predicciones de probabilidad sesgadas y algo parecido ocurre con los tests de contrastación de hipótesis.

Para evitar estos problemas, deben utilizarse otras especificaciones para un modelo de este tipo. Algunas posibles soluciones han sido expuestas en este artículo, de todas ellas la del modelo logístico es preferible a cualquiera otra. Las técnicas de estimación de datos agrupados, basadas en la formulación de modelos no lineales ex-post o ex-ante, presentan el doble inconveniente de reducir la información, al tiempo que debilitan el rigor econométrico en la medida en que los problemas de transformación de un modelo no lineal en uno lineal son prácticamente eludidos. El modelo logístico, en cambio, permite evitar esos dos inconvenientes y presenta la ventaja de proporcionar mejores predicciones de la probabilidad de realización del suceso explicado. El problema que presenta tal modelo es el de su estimación, ya que debe emplearse para ello un proceso de optimización numérica, lo que no es simple; pero, siendo convexa la función de verosimilitud obtenida, existe un máximo global y los estimadores de los parámetros poseen todas las propiedades de los estimadores maximoverosímiles, esto es, consistencia, eficiencia asintótica y normalidad asintótica.

La generalización del modelo logístico de variable dependiente binaria al de variable múltiple permite, por su parte, tratar situaciones más variadas y consiguientemente más adaptadas a la realidad, aunque a su estimación, además de los problemas que acabamos de enumerar en el modelo precedente, hay que añadir un mayor número de parámetros a estimar para un mismo número de variables exógenas.

Por último, se ha presentado el sistema de ecuaciones logísticas simultáneas de variables endógenas cualitativas, que permitirá formalizar situaciones muy frecuentes en economía, que deben estimarse conjuntamente, ya que tratar separadamente las ecuaciones que componen tal sistema y estimarlas aisladamente, sesgaría el valor de los estimadores de los parámetros, de igual forma que ocurre en el sistema de ecuaciones simultáneas de variables endógenas continuas.

La utilización de un modelo logístico estimado por el método de la máxima verosimilitud se revela, así, como el más apropiado para la estimación de modelos de variables dependientes cualitativas, tanto en el caso de modelos de elección binaria, como múltiples e incluso en el caso de sistemas de ecuaciones simultáneas.

BIBLIOGRAFIA

AIGNER, D. S.; GOLBERGER, A. S.: KALTON, G.: "On the Explanatory Power of Dummy Variable Regressions". *International Economic Review,* 16, 1975, pp. 505-510.

AITCHISON, J; BENNETT, J.: "Polychotomous Quantal Reponse by Maximun Indicant". *Biometrika,* 57; 1970, pp. 253-262.

AMEMIYA, T.: "Regression Analysis When the Dependent Variable is Truncated Normal". *Econometrica,* 41; 1973, pp. 997-1.016.

AMEMIYA, T.: "Qualitative Reponse Models". *Annals of Economic and Social Measurement,* 4; 1975, pp. 363-372.

AMEMIYA, T.: "The Estimation of a Simultaneous Equation Generalized Probit Model". *Econometrica,* 46, 1978, pp. 1.193-1.205.

AMEMIYA, T; NOLD, F.: "A Modified Logit Model", *Review of Economics and Statistics,* 57, 1975, pp. 255-257.

ASHFORD, J. R.; SOWDEN, R. R.: "Multivariate Probit Analysis". *Biometrics,* 26; 1970, pp. 535-546.

ASHTON, W. D.: *"The Logit Transformation".* Hafner, New York, 1972.

BERKSON, J.: "Applications of the Logistic Function to Bioassay". *Journal of the American Statistical Association,* 39; 1949, pp. 357-365.

BERKSON, J.: "Maximum Likelihood and Minimum Chi-Square Estimations of the Logistic Finction". *Journal of American Statistical Association,* 50; 1955, pp. 130-161.

COX, D. R.: "The Analysis of Multivariate Binary Data". *Applied Statistics,* 21; 1972, pp. 113-120.

CRAGG, J. G.: "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods". *Econometrica,* 39; 1971, pp. 829-844.

EFRON, B.: "The Efficiency of Logistic Regression Composed to Normal Discriminant Analysis". *Journal of American Statistical Association,* 70; 1975, pp. 892-898.

FINNEY, D.: *"Probit Analysis".* New York: Cambridge University Press, 1971.

FRIEDMAN, P.: "Suggestions for the Analysis of Qualitative Dependent Variables". *Public Finance Quaterly,* 1; 1973, pp. 345-355.

GOLDBERGER, A. S.: *"Econometric Theory".* Wiley, New Yor, 1972.

GOLDFELD, S; QUANDT, R. E.: *"Nonlinear Methods on Econometrics".* Amsterdam; North Holland Publishing Company, 1972.

GOODMAN, L. A.: "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications". *Journal of the American Statistical Association,* 65; 1970, pp. 226-256.

GOODMAN, L. A.: "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables". *American Sociological Review,* 37; 1972, pp. 28-46.

GRIZZLE, J. E.: "Multivariate Logit Analysis". *Biometrics,* 27; 1971, pp. 1.057-1.062.

GUNDERSON, M.: "Retention of Trainees, a Study with Dichotomous Dependent Variables". *Journal of Econometrics,* 2; 1974, pp. 79-93.

HAUSMAN, J. A.; WISE, D. A.: "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences". *Econometrica,* 46; 1978, pp. 403-426.

HECKMAN, J. J.: "Dummy Exogenous Variables in a Simultaneous Equations System". *Económetrica,* 46; 1978, pp. 931-959.

HULETT, J. R.: "On the Use of Regression Analysis with a Qualitative Dependent Variable". *Public Finance Quaterly,* 1; 1973, pp. 339-344.

LASSIBILLE, G.: "L' Estimation de Modèles à Variable Dependente Dichotomique" *Revue d' Economie Appliquée,* 32; 1979, pp. 375-393.

LEE, L. F.: "Identification and Estimation in Binary Choice Models with Limited (censored) Dependent Variables". *Econometrica,* 47; 1979, pp. 977-996.

LEE, L. F.; TROST, R. P.: "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand". *Journal of Econometrics,* 8; 1978, pp. 357-382.

LI, M.: "A Logit Model of Home Ownership". *Econometrica,* 45; 1977, pp. 1.081-1.097.

MADDALA, G. S.: "Selectivity Problems in Longitudinal Data". *Annals de l' INSEE,* 1978, pp. 421-450.

MADDALA, G. S.; LEE, L. F.: "Recursive Models with Qualitative Endogeneous Variables". *Annals of Economics and Social Measurement",* 5; 1976.

MADDALA, G. S.; TROST, R.: "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand". *Journal of Econometrics,* 8; 1978, pp. 357-382.

MANSKI, C.; LERMAN, S.: "The Estimation of Choice Probabilities from Choice Based Samples". *Econometrica*, 45; 1977, pp. 1.977-1.988.

McFADDEN, D.: "Quantal Choice Analysis: A survey". *Annals of Economic and Social Measurement*, 5; 1976, pp. 363-390.

McFADDEN, D.: "A Comment on Discriminant Analysis 'versus' Logit Analysis". *Annals of Economics and Social Measurement*, 5; 1976, pp. 511-523.

McGILLIVRAY, R. G.: "Estimating the Linear Probability Function". *Econometrica*, 38; 1970, pp. 775-776.

McGILLIVRAY, R. G.: "Binary Choice of Urban Transport Mode in the San Francisco Bay Region". *Econometrica*, 40; 1972, pp. 827-848.

NERLOVE, M.: "Econometric Analysis of Longitudinal Data: Approaches, Problems and Prospects". The Econometrics of Panel Data. *Annals de l'INSEE*, 1978, pp. 30-32, 7-22.

NERLOVE, M.; PRESS, J.: "Univariate and Multivariate Log-Linear and Logistic Models". *RAND Report'* No. R-1036- EDA/NIH, Santa Mónica, California, 1973.

PRAIS; S. J.; HOUTHAKKER, H. S.: *'The Analysis of Family Budgets"*. Cambridge University Press, 1955.

SCHMIDT, P.; STRAUSS, R. P.: "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach". *Econometrica*, 43; 1975, pp. 745-755.

THEIL, H.: "A Multinomial Extension of the Linear Logit Model". *International Economic Review*, 10; 1969, pp. 251-259.

THEIL, H.: "On the Estimation of Relationships Involving Qualitative Variables". *American Journal of Sociology*, 76; 1970, pp. 103-154.

THEIL, H.: *"Principles of Econometrics"*. Wiley, New York 1971.

WALKER, S.; DUNCAND, D.: "Estimation of the Probability of an Event as a Function of Several Independent Variables". *Biometrika*, 54; 1967, pp. 167-179.

ZELLNER, A.; LEE, T.: "Joint Estimation of Relationships Involving Discrete Random Variables". *Econometrica*, 33; pp. 383-394.

LOS NUEVOS PLANTEAMIENTOS DE LA CIENCIA ECONOMICA ANTE LA CRISIS

Antonio García Lizana
Guillermina Martín Reyes

Sumario: 1. Keynes y el lado de la demanda.- 2. La apoteosis de la política económica keynesiana.- 3. Se inicia el declive.- 4. Hechos, interrogantes y teorías.- 5. Las cuatro líneas de aproximación.- 6. El contenido de la Economía del lado de la oferta.- 7. La "reconstrucción" de la Macroeconomía.- 8. La administración de la oferta y la rebelión fiscal.- 9. Pensando en España.

El pensamiento económico concedió en otro tiempo una relevante importancia al tema de la oferta. Basta pensar que de acuerdo con la ley de Say no es el gasto quien determina la producción, sino que es ésta la que determina la renta y el gasto. No es la demanda quien da lugar y estimula a la oferta, sino que ocurre exactamente lo contrario.

Pero con la crisis de los años 30 cambiaron muchas cosas en el pensamiento económico. La amplitud y gravedad de la Gran Depresión estimuló la atención, la inteligencia y la imaginación de economistas y políticos en la búsqueda de teorías capaces de explicar lo que estaba ocurriendo, y en la aplicación de nuevas soluciones, puesto que las tradicionales (las "clásicas") parecían no ser las adecuadas. Esto supuso, entre otras cosas, que se estableciera una muy marcada separación entre el campo de la Microeconomía y el de la Macroeconomía. En realidad, la Macroeconomía, tal como hoy la conocemos, nació en el seno de la Gran Depresión.

Con la Gran Depresión se reforzó la idea de que la economía privada, tomada en su conjunto, era en sí misma inestable, y que, correspondientemente, los gobiernos debían asumir una mayor responsabilidad a fin de lograr la estabilización económica nacional. Por otra parte, dicha falta de estabilidad estaba —se entendió— muy