IV Congreso de Economía Regional de Castilla y León **Burgos 1994, Comunicaciones 2:** 808-821

PREDICCIONES BAYESIANAS DE LAS PROBABILIDADES DE CRECIMIENTO DEL EMPLEO REGIONAL.

José Luis ROJO GARCÍA
José Antonio SANZ GÓMEZ
Universidad de Valladolid
Dpto. Economía Aplicada (Estadística y Econometría)

1. INTRODUCCIÓN

En el campo de la estimación de magnitudes económicas, los analistas nos planteamos con frecuencia la evaluación de las posibilidades que tienen de presentarse ciertos sucesos de interés. El amplio desarrollo de las aplicaciones estadísticas y econométricas en este ámbito hace que la modelización de estos fenómenos revista una naturaleza estadística, por lo que dichas posibilidades se evalúan mediante el cálculo de las probabilidades de dichos sucesos dentro de un modelo coherente con la realidad que se trata de describir.

Los sucesos en los que intervienen **umbrales** son relativamente frecuentes. Dada una magnitud, un umbral es una cantidad tal que, si la magnitud lo rebasa, se produce un determinado efecto. El suceso consistente en la presencia de dicho efecto resulta, por tanto, equivalente. a la superación del umbral por dicha magnitud. La probabilidad de ocurrencia de dicho suceso puede, en definitiva, interpretarse como la de que la magnitud rebase el umbral.

Los ejemplos de estos problemas en la literatura económica, desde un enfoque econométrico, son abundantes y se presentan cuando la magnitud en estudio se explica por otras, en el marco de un modelo econométrico cuya variable endógena es la citada magnitud. Sin pretender ser exhaustivos, citemos a HECKMAN Y MACURDY (1980) **(1)**, FASE (1971) **(2)**, ASHFORD Y SOWDEN (1970) **(3)**, ZELLNER, HONG Y GULATI (1990) **(4)**, o ROJO (1990) **(5)**.

En los primeros, la magnitud no es observable, por lo que la solución pasa por construir una variable dicotómica que vale la unidad cuando se produce el suceso, es decir, cuando la magnitud rebasa el umbral. Así, llamando Y a la magnitud, y denotando por a al umbral, la situación se describe esquemáticamente como

$$\begin{array}{lll} \text{Si } Y > a & \text{definimos} & Y^* = 1 \\ \text{Si } Y \leq a & \text{definimos} & Y^* = 0 \end{array}$$

En los dos últimos, la magnitud es observable en el periodo histórico, por lo que la construcción de una variable dicotómica, que es la técnica habitualmente utilizada, produce una importante pérdida de información. Ésta, como ilustramos en la aplicación, es especialmente relevante en problemas temporales, lo que desaconseja su conversión en modelos de respuesta cualitativa.

El trabajo que aquí se presenta es una aplicación a la economía castellano-leonesa de la técnica desarrollada por los autores, que permite predecir, de forma óptima con un cierto criterio, la probabilidad de que la variable endógena de un modelo lineal clásico supere un umbral prefijado. En concreto, predeciremos la probabilidad de que el empleo crezca en nuestra Comunidad Autónoma durante el año 1994 y durante 1995. Esta técnica tiene aún un alcance limitado ya que, normalmente, las ecuaciones que explican el comportamiento económico tienen una naturaleza simultánea, por lo que habitualmente cada ecuación incluye variables endógenas entre las explicativas. Por ello, se trata de un avance que esperamos mejorar más adelante.

La técnica utilizada recibe el nombre de bayesiana, porque la naturaleza óptima de la probabilidad que se estima deriva del hecho de que es la estrategia de Bayes en un problema de decisión, con una función de pérdida adecuada al mismo.

2. OBTENCIÓN DEL ESTIMADOR ÓPTIMO

Sea $Y = (Y_1, \ldots, Y_T)'$ la matriz que contiene los valores muestrales de la magnitud cuyo comportamiento estamos estudiando. Se supone que dicha variable se explica a través de k variables predeterminadas, cuyos valores muestrales describiremos como $X_1, \ldots, X_k$, con $X_i = (x_{1i}, \ldots, x_{Ti})'$. Llamando X a la matriz $T \times k$ que agrupa los valores muestrales de dichas variables, esto es, a la matriz $X = (X_1, \ldots, X_k)$, suponemos que podemos describir su relación mediante el modelo econométrico uniecuacional $Y = X\beta + \epsilon$, verificando las hipótesis del modelo lineal clásico (MLC):

1. La matriz $\mathbf{X}$ es de rango k y no aleatoria.
2. Las perturbaciones, $\epsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, son v.a.i.i.d., siendo su distribución $N(0, \sigma)$.
3. Existe el límite $\lim_{T\to\infty} \frac{X'X}{T} = Q$, y Q es una matriz cuadrada de orden k, no singular.

Además, supondremos en nuestro trabajo que la matriz de los k parámetros β, es una matriz **aleatoria** cuyos elementos son independientes de ϵ.

Denotaremos por X^0 al vector fila de los valores de las variables exógenas en el periodo de predicción, $X^0 = (x_1^0, \ldots, x_k^0)$. En este periodo el modelo será, por tanto, $Y^0 = X^0\beta + \varepsilon^0$, donde también supondremos que la perturbación ε^0 es $N(0, \sigma)$ y está incorrelacionada con las perturbaciones correspondientes al periodo muestral siendo, asimismo, independiente de β.

Sea a un cierto umbral para la variable endógena. Pretendemos estimar de forma óptima la probabilidad de que la endógena en el periodo de predicción, Y^0, rebase el umbral, es decir, estimar $p_\beta = p(Y^0 > a \mid \beta)$.

Considerada la función de pérdida cuadrática $L(\beta, \tilde{\beta}) = [p_\beta - p_{\tilde{\beta}}]^2$, donde p_β y $p_{\tilde{\beta}}$ denotan la probabilidad de superación del umbral por la variable endógena en el periodo de predicción bajo cualquier parámetro β y bajo una especificación $\tilde{\beta}$ del mismo, respectivamente, ROJO Y MARTÍN (1988) suponen una distribución neutral *a priori* para los parámetros y obtienen la condición que debe cumplir el predictor óptimo, $X^0\tilde{\beta}$, de la variable endógena cuando σ es conocida, resultando que $\tilde{\beta}$ ha de ser solución de la igualdad,

$$\left(\frac{|X'X|}{e|Z'Z|}\right)^{\frac{1}{2}} \cdot p\left(N(0,1) \geq \frac{\sqrt{e}}{\sigma}(a - X^0\hat{\beta})\right) = p\left(N(0,1) \geq \frac{a - X^0\tilde{\beta}}{\sigma}\right),$$

siendo e el escalar

$$e = 1 - X^0(X'X + X^{0'}X^0)^{-1}X^{0'} = \frac{1}{1 + X^0(X'X)^{-1}X^{0'}}.$$

En SANZ (1994) hemos demostrado que $\frac{|X'X|}{e|Z'Z|}$ vale la unidad, por lo que la expresión anterior puede reducirse y obtenerse explícitamente una predicción bayesiana,

$$X^0\tilde{\beta} = a(1 - \sqrt{e}) + \sqrt{e}X^0\hat{\beta} \ ,$$

que resulta ser una combinación lineal convexa del umbral a y del predictor mínimo cuadrático, $X^0\hat{\beta}$.

En definitiva, la probabilidad óptima de superación del umbral bajo el predictor bayesiano, por parte de la variable endógena resulta

$$p_{\tilde{\beta}} = p(Y^0 > a \mid \tilde{\beta}) = 1 - \Phi\left(\sqrt{e} \cdot \frac{a - X^0\hat{\beta}}{\sigma}\right) .$$

Denotando, en el mismo sentido que antes, por $p_{\hat{\beta}}$ a la probabilidad de superación del umbral por la variable endógena en el periodo de predicción bajo el predictor MCO, $\hat{\beta}$, resulta

$$p_{\hat{\beta}} = p(Y^0 > a \mid \hat{\beta}) = 1 - \Phi\left(\frac{a - X^0\hat{\beta}}{\sigma}\right) .$$

Si la varianza de la perturbación, σ^2, es desconocida, hemos demostrado (SANZ (1994)) que los valores ($\tilde{\beta}$ y $\tilde{\sigma}^2$) que minimizan la función de riesgo, con una distribución *a priori* neutral, son aquéllos que cumplen

$$p(t_\nu) = p\left(t_\nu \geq \sqrt{\frac{e}{S^2}}(a - X^0\hat{\beta})\right) = p\left(N(0,1) \geq \frac{a - X^0\tilde{\beta}}{\tilde{\sigma}}\right) .$$

Resulta en este caso, no una predicción óptima, sino una cierta "tipificación" de la diferencia entre ella y el umbral. Es decir, a lo sumo obtendremos $\frac{a-X^0\tilde{\beta}}{\tilde{\sigma}}$. La probabilidad óptima, $p(t_\nu)$, que aparece a la izquierda de la expresión anterior se obtiene, por tanto, como

$$p(t_\nu) = 1 - F_{t_\nu}\left(\frac{a - X^0\hat{\beta}}{S/\sqrt{e}}\right) ,$$

es decir, la función de distribución de una variable aleatoria t de Student con $\nu = T - k$ grados de libertad, calculada en el punto indicado.

La probabilidad que proporcionaría la aplicación de los métodos clásicos sería

$$p_{\hat{\beta},\hat{\sigma}^2} = p(N(X^0\hat{\beta}, \hat{\sigma}) \geq a) = 1 - \Phi\left(\frac{a - X^0\hat{\beta}}{\hat{\sigma}}\right),$$

donde $\hat{\beta}$ es el estimador mínimo cuadrático de β, y $\hat{\sigma}^2 = S^2$ es el estimador insesgado clásico de σ^2 obtenido a partir de los residuos mínimo cuadráticos.

La diferencia entre ambas probabilidades estimadas es que sometemos, en ambos casos, al umbral, a, al mismo "cambio de origen", $X^0\hat{\beta}$, aunque a distinto "cambio de escala", $\frac{S}{\sqrt{e}}$ o S, siendo el primero siempre mayor que el segundo (ya que e es siempre menor que la unidad).

A partir de las anteriores expresiones se comprueban los siguientes resultados (SANZ (1994)), que relacionan el predictor obtenido con el de mínimos cuadrados ordinarios.

1. Cuando σ es conocida, el predictor bayesiano de la variable endógena, $X^0\tilde{\beta}$, se encuentra siempre entre el predictor clásico, $X^0\hat{\beta}$, y el umbral, a. Esto es, la función de pérdida *atrae* la predicción hacia el umbral.

2. En consecuencia, y tanto si la varianza de las perturbaciones es conocida como si no lo es, la estimación óptima de la probabilidad está más *equilibrada* que la clásica. En concreto, cuando σ es conocida, si $p_{\hat{\beta}} < \frac{1}{2}$, entonces $p_{\hat{\beta}} < p_{\tilde{\beta}} < \frac{1}{2}$; por contra, si $p_{\hat{\beta}} > \frac{1}{2}$, entonces $p_{\hat{\beta}} > p_{\tilde{\beta}} > \frac{1}{2}$.

   Cuando σ es desconocida, $p_{\hat{\beta},\hat{\sigma}^2} < p_{\tilde{\beta},\tilde{\sigma}^2} < \frac{1}{2}$ si $p_{\hat{\beta},\hat{\sigma}^2} < \frac{1}{2}$, cambiando las desigualdades si $p_{\hat{\beta},\hat{\sigma}^2} > \frac{1}{2}$.

3. La distancia absoluta entre ambas predicciones (y, por tanto, entre ambas probabilidades), se hace tanto menor cuanto más próxima esté la predicción clásica, $X^0\hat{\beta}$, al umbral, a.

4. La distancia relativa entre ambas predicciones, esto es, el cociente

$$\frac{|X^0\tilde{\beta} - X^0\hat{\beta}|}{|a - X^0\hat{\beta}|}$$

depende del valor de e y, en consecuencia, del de la forma cuadrática $X^0(X'X)^{-1}X^{0'}$, de forma que, para los modelos con término constante (esto es, $X_1 = (1, \ldots, 1)'$) se hace mínima para $X^0 = (1, \overline{X}_2, \ldots, \overline{X}_k)$, donde $\overline{X}_i$ es la media aritmética de los valores muestrales de la variable X_i. Para una especificación concreta de X^0, la proximidad relativa entre ambas predicciones (y, por tanto, recordemos, entre ambas probabilidades), depende del aspecto de la forma cuadrática dada por la matriz $(X'X)^{-1}$.

3. UNA APLICACIÓN. COMPORTAMIENTO DEL EMPLEO REGIONAL A CORTO PLAZO

Vamos a aplicar la técnica anterior al análisis de las probabilidades de crecimiento del empleo en Castilla y León en 1994 y 1995. Como decíamos en la introducción, la técnica desarrollada por nosotros es aún muy limitada, y no puede utilizarse en estos momentos más que para modelos uniecuacionales sencillos. Dentro de nuestro Departamento, hemos formado un pequeño grupo de trabajo que actualmente explora las posibilidades de resolver el mismo problema para modelos uniecuacionales más complicados (y, por tanto, más realistas), e incluso para modelos de ecuaciones simultáneas.

Por todo ello, los resultados que a continuación se presentan tienen más valor como ilustración de las posibilidades del método que como una aplicación plenamente desarrollada.

El equipo HISPALINK-CL, del que los autores son integrantes, elabora predicciones de los valores añadidos brutos a precios de mercado y del empleo (media anual de ocupados) con una desagregación a nueve ramas de la economía regional: Agricultura, cuatro ramas industriales (Energía, Bienes de equipo, Bienes intermedios y Bienes de consumo), Construcción y tres ramas de Servicios (Transportes y Comunicaciones, Servicios destinados a la venta y Servicios no destinados a la venta). Asimismo, realiza predicciones para la renta y el consumo regional, y estudia la evolución futura del desempleo en nuestra Comunidad.

La predicción del empleo regional se realiza, en el seno del modelo, con la misma desagregación, en un bloque multiecuacional en el que sus valores para cada rama se hacen depender, en general, de variables endógenas ligadas a la producción (en la forma de funciones de producción despejadas) y, a veces, del comportamiento de variables nacionales como variables vicarias de una tendencia general. No resulta posible, por tanto, formular una única ecuación de comportamiento con capacidad explicativa del empleo en su conjunto, ya que las pautas que rigen, por ejemplo, la evolución del empleo agrario, no pueden armonizarse con las que determinan dicha magnitud en las ramas industriales.

Nuestra aplicación propone una única ecuación para explicar el empleo total regional, de la forma

$$ET = f(VT, EET) ,$$

donde ET es el empleo regional, EET el nacional (ambos medidos como media anual de ocupados, en miles) y VT el valor añadido bruto a p.m. regional, en millones de pesetas de 1986. Los datos de empleo provienen de E.P.A. hasta 1993, y los nacionales de 1994 y 1995 son predicciones del Centro L.R. Klein. VT está tomado del banco de datos HISPADAT, que extrae su información de la Contabilidad Regional de España desde 1980 hasta 1992, del banco de datos HISPALINK entre 1972 y 1979, y de predicciones del grupo HISPALINK-CyL para 1994 y 1995 (obviamente, todos los valores están debidamente armonizados).

Como puede verse, el modelo es muy simple como para tener potencia explicativa; además, aunque puede aceptarse la exogeneidad de EET, ya que, al ser una variable nacional puede suponerse que viene dada desde el entorno del sistema (podría discutirse esta afirmación ya que, por un lado, ET es uno de los sumandos cuya agregación proporciona EET, y, por otro, el empleo regional determina parcialmente la demanda regional que se satisface por la producción generada por los ocupados nacionales, EET; no obstante, puede aceptarse que las relaciones son, en este sentido, muy débiles), en ningún caso es aceptable la exogeneidad de VT. En definitiva, la predicción del empleo regional en 1994, desde el punto de vista estadístico, estará condicionada por el valor añadido regional de 1994.

En el año 1993, el empleo en Castilla y León ascendió a 783.7 miles de personas (EPA, media anual). Nuestro objetivo es la estimación de la probabilidad de que el empleo regional crezca en 1994, es decir, que la media anual de ocupados en este año supere el valor de 1993, esto es, que supere los 783.7 miles de ocupados. En este caso, el modelo a estimar es

$$ET_t = \beta_0 + \beta_1 \cdot VT_t + \beta_2 \cdot EET_t + \varepsilon_t \ ,$$

los datos muestrales son las 17 observaciones de cada variable correspondientes al periodo 1977-1993 que se presentan en la tabla 1 al final de este trabajo, y el problema que tratamos de resolver es la estimación de la probabilidad de que ET_{1994} (Y^0 con la notación del apartado anterior) supere el umbral de 783.7.

CUADRO 1

ESTIMACION DE LA ECUACION DE COMPORTAMIENTO DEL EMPLEO REGIONAL

LS // Dependent Variable is ET
SMPL range: 1977 - 1993
Number of observations: 17

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	411.02655	29.544141	13.912286	0.0000
VT	-0.0001436	9.846E-06	-14.585214	0.0000
EET	0.0578713	0.0027700	20.892385	0.0000

R-squared	0.971310	Mean of dependent var	813.6408
Adjusted R-squared	0.967211	S.D. of dependent var	35.78825
S.E. of regression	6.480417	Sum of squared resid	587.9412
Log likelihood	54.24097	F-statistic	236.9861
Durbin-Watson stat	1.741749	Prob(F-statistic)	0.000000

El cuadro 1 muestra la estimación de la ecuación de comportamiento. Puede observarse en él la deficiente capacidad explicativa del modelo adoptado; en efecto, el coeficiente correspondiente al valor añadido regional es negativo, cuando no debiera ocurrir así. El motivo de este hecho reside en que el empleo nacional recoge prácticamente toda la variación de ET, pero muestra crecimientos superiores a los de esta magnitud, como puede observarse en el cuadro 2; el efecto del valor añadido regional en la ecuación es únicamente el de descontar una parte del crecimiento que induciría el empleo nacional en las predicciones del regional. En cualquier caso, el fondo del problema es que la especificación de la ecuación no es la adecuada, como indicamos anteriormente.

CUADRO 2
VARIACION DEL EMPLEO REGIONAL Y NACIONAL
Y DEL V.A. REGIONAL POR PERIODOS
(Tasa anual acumulativa, en %)

	ET	EET	VT
1977-1985	-1.9	-1.82	1.29
1985-1991	1.97	2.87	2.11
1991-1993	-3.5	-3.11	0.39

De la anterior ecuación obtenemos $\hat{\beta}$ y el error estándar, $\hat{\sigma}$. El vector de las variables predeterminadas en el periodo de predicción es

$$X^0 = (1\ ,2170217\ ,11719)\ ,$$

que corresponden al término constante, al valor añadido regional en 1994 y al empleo nacional en dicho periodo (véase la penúltima fila de la tabla 1).

Con esto, es inmediato obtener, a la vista de lo expuesto en el apartado anterior, la probabilidad óptima que buscamos. Así, la predicción con los métodos clásicos del empleo regional es $X^0\hat{\beta} = 777.58$, el umbral es $a = 783.7$, y el valor de e resulta $e = 0.847209$. Por tanto, la predicción bayesiana del empleo regional en 1994 resulta

$$X^0\tilde{\beta} = 778.064\ .$$

En definitiva, nuestra predicción eleva la cifra en cerca de quinientos ocupados. Como dijimos en el apartado anterior, *atrae* la predicción MCO hacia el umbral por lo que, como éste es superior a aquélla, eleva el valor de la predicción, si bien levemente, ya que $\sqrt{e} = 0.92044$ está próximo a la unidad.

Obtengamos las predicciones de la probabilidad de crecimiento del empleo. En el apartado anterior demostramos que la predicción óptima era

$$p_{\tilde{\beta}} = 1 - \Phi\left(\sqrt{e} \cdot \frac{a - X^0\hat{\beta}}{\sigma}\right),$$

por lo que, estimando σ como $S = 6.4804$, (6), y sustituyendo el resto de los valores resulta una probabilidad $p_{\tilde{\beta}} = 0.19224$.

En idéntico sentido, podemos calcular la probabilidad de superar el umbral utilizando los métodos clásicos,

$$p_{\hat{\beta}} = 1 - \Phi\left(\frac{a - X^0\hat{\beta}}{\sigma}\right) ;$$

sustituyendo las expresiones correspondientes resulta, $p_{\hat{\beta}} = 0.17237$.

En definitiva, con la aplicación de este procedimiento óptimo, la probabilidad de crecimiento del empleo regional aumenta dos puntos porcentuales. Con estos datos, puede plantearse el lector estimar otras probabilidades. Por ejemplo, la de que el empleo no decrezca por debajo del 1%, esto es, que el empleo regional en 1994 sea superior a 775.863 miles de ocupados, resultaría, con la predicción MCO (que no cambia), una predicción bayesiana del empleo regional de 777.441, que reduce en poco mas de un centenar el número de ocupados (*atrae* la predicción MCO hacia el umbral). Además, las predicciones óptimas de la probabilidad resultan, con los métodos clásicos,

$$p_{\hat{\beta}} = p(ET_{1994} > 775.863|\hat{\beta}) = 0.60431$$

y, con la nuestra,

$$p_{\tilde{\beta}} = p(ET_{1994} > 775.863|\tilde{\beta}) = 0.59618 .$$

Ahora, la aplicación de nuestro procedimiento, mantiene ligeramente por debajo nuestra predicción. Este hecho es debido, como apuntamos en el apartado anterior, a la mayor proximidad del umbral a la predicción mínimo cuadrática (y, por tanto, entre ambas predicciones), lo que conlleva una mayor aproximación de las probabilidades predichas.

Obsérvese, finalmente, que la utilización de los modelos de respuesta cualitativa en este problema suponen una importante pérdida de información. Si construimos la variable dicotómica

$$\begin{array}{lll} \text{Si } ET > 783.7 & \text{definimos} & Y^* = 1 \\ \text{Si } ET \leq 783.7 & \text{definimos} & Y^* = 0 \end{array}$$

perdemos la información referente a la diferencia entre los valores muestrales de ET y el umbral.

Podemos, finalmente, plantearnos predecir la probabilidad de que, ya en 1995, se produzca un crecimiento del empleo regional. Como el valor de 1994 no es conocido, sino que está predicho, y nuestro procedimiento no permite aún aplicarse a varios periodos, la predicción estará condicionada por el valor de 1994 que, con nuestra técnica, resultó de 778.065 miles de ocupados. En la última fila de la tabla 1 se presentan los valores de las variables predeterminadas. Realizando cálculos análogos a los anteriores, se obtienen las predicciones óptimas de la probabilidad:

$$p_{\hat{\beta}} = 0.35677 \quad \text{y} \quad p_{\tilde{\beta}} = 0.36789 \; ,$$

aumentando un punto porcentual la predicción óptima de probabilidad, siendo, por otra parte, ambas predicciones muy similares ($X^0\hat{\beta} = 775.765$ y $X^0\tilde{\beta} = 775.951$), es decir, predecimos cerca de doscientos empleos más con nuestra técnica que con la clásica (de nuevo, como decíamos, *atrae* la predicción MCO hacia el umbral), estando ligeramente por encima la predicción de la probabilidad de que el empleo crezca en 1995.

4. CONCLUSIONES

En este trabajo predecimos la probabilidad de que crezca el empleo regional en los años 1994 y 1995. Para ello, aplicamos una técnica original, que permite construir un predictor óptimo de la probabilidad de que la variable endógena de un modelo lineal clásico rebase un umbral prefijado. Este predictor resulta ser un punto intermedio entre el mínimo cuadrático y el umbral, y la proximidad relativa entre ambos predictores depende del comportamiento de una cierta forma cuadrática ligada a los valores muestrales de las variables predeterminadas y de los valores de éstas en el periodo de predicción.

En definitiva, la probabilidad de que el empleo crezca en nuestra Comunidad, resulta algo superior con este procedimiento óptimo a la que se obtendría con los procedimientos clásicos. Otro tanto ocurre para el año 1995. En cualquier caso, las diferencias entre ambas probabilidades no son elevadas, debido a la proximidad del umbral a la predicción del empleo con el método clásico.

TABLA 1

VALORES MUESTRALES DE LAS VARIABLES

	EET	VT	ET
1977	12328	1699074	872.78
1978	12103	1715636	864.25
1979	11912	1689666	855.40
1980	11557	1715389	841.25
1981	11231	1682637	818.57
1982	11117	1727381	806.48
1983	11044	1791144	800.92
1984	10743	1831297	778.05
1985	10641	1882830	748.80
1986	10881	1903944	759.52
1987	11469	1987605	790.57
1988	11773	2064761	797.27
1989	12258	2100650	816.82
1990	12579	2118510	841.97
1991	12609	2134713	841.60
1992	12368	2114782	813.92
1993	11837	2151279	783.70
1994	11719	2170217	—
1995	11813	2220615	—

NOTAS

(1) Los autores estudian la diferencia, Y_{it}, para la i-ésima mujer en el periodo t, entre el salario de reserva y el ofertado por el mercado de trabajo. Si la diferencia es positiva, esto es, si $Y_{it} > 0$, la mujer decide continuar fuera de la fuerza laboral en dicho periodo. El umbral sería igual a cero.

(2) Clasifica el autor a un individuo i como pobre en el periodo t si su renta, $E(i,t)$, no supera un valor, C, constante. El umbral sería la cantidad C o, si la variable es $Y_{it} = E(i,t) - C$, el umbral sería cero.

(3) Un minero de carbón presenta problemas respiratorios cuando su tolerancia, y_i^* no rebasa una constante, γ, desconocida. Se trataría de estimar la probabilidad de que tenga problemas respiratorios, esto es, que $y_i^* < \gamma$.

(4) Los autores analizan la probabilidad de que el P.I.B. presente un punto de retorno en un determinado periodo, esto es, un punto en el que se cambie de crecimiento a decrecimiento o viceversa. Un punto, t, de retorno hacia arriba sería aquél en el que, dado que $PIB_{t-1} > PIB_t$, ocurre que $PIB_{t+1} > PIB_t$. El problema puede estudiarse condicionado por los valores anteriores y contemporáneos a t, y tomar como umbral el valor PIB_t, siendo la variable endógena PIB_{t+1}.

(5) El autor estudia la probabilidad, en un modelo econométrico, de que el crecimiento regional supere al nacional, de que aumente la población ocupada o de que el desempleo se reduzca en un año concreto.

(6) El valor de σ es desconocido, por lo que lo sustituimos por una estimación suya para conseguir que el predictor sea identificable.

REFERENCIAS

ASHFORD, J. R. Y SOWDEN, R. R. (1970). Multivariante Probit Analysis. Biometrics, **26**, 535-546.

FASE, M. M. G. (1971). On the estimation of lifetime income, J.A.S.A., vol. 66, **366**, 686-692.

HECKMAN, J. J. Y MACURDY, T. E. (1980). A Life Cycle Model of Female Labour Supply, Review of Economic Studies, **47**, 47-74.

ROJO, J. L. Y MARTÍN, I. (1987). Predicciones de variables endógenas en problemas ligados a la existencia de umbrales. Estadística Española, (I.N.E.), **116**, 45-53.

ROJO, J. L. (1990). Trabajo de Investigación (no publicado), para el concurso al cuerpo de Catedráticos de Universidad (Area de Economía Aplicada).

SANZ, J. A. (1994). Predicciones bayesianas de probabilidad en presencia de umbrales en un modelo lineal. Una aplicación a la predicción del gasto turístico. Tesis doctoral, Departamento de Economía Aplicada (Estadística y Econometría), Universidad de Valladolid.

ZELLNER, A.; HONG, C. Y MITU GULATI, G. (1990). Turning points in Economic Time Series, Loss structure and Bayesian forecasting. *Bayesian and Likelihood Methods in Statistics and Econometrics.* Essays in Honor of George A. Barnard. Ed. Geisser, S.; Hodges, J. S.; James Press, S. y Zellner, A., pp. 371-393. Amsterdam: North-Holland.