ITEA (1999), Vol. 95V N.° 1, 21-29

DETERMINACIÓN DE INDICADORES DE AGRICULTORES CON RENTA AGRARIA EXCLUSIVA EN ANDALUCÍA

M. Arriaza
P. Ruiz

Centro de Investigación y Formación Agraria
Consejería de Agricultura y Pesca
Junta de Andalucía
Apartado 3092
14080 Córdoba
España

RESUMEN

En el presente trabajo se determinan los indicadores socioeconómicos que caracterizan a los agricultores no ganaderos andaluces cuya única fuente de renta proviene de la actividad agraria. Basado en una encuesta realizada en Andalucía en verano de 1996, el modelo logístico seleccionado muestra como el nivel educativo y el tiempo que el agricultor dedica a la actividad agraria son las variables de mayor relevancia, mientras que las variables relacionadas con la orientación técnica productiva parecen no contribuir a explicar la variable renta.

Palabras Clave: Modelo logit, Variable ficticia, Renta agraria.

SUMMARY
SOCIO-ECONOMIC INDICATORS OF FARMERS WITH INCOME ONLY FROM AGRARIAN ACTIVITIES

This paper presents an econometric research on limited-dependent variable models, in this case a dummy dependent variable which represent the probability of encountering farmer which income comes only from farming, provided the values of the regressors. A logit model is applied to the data from a survey on Andalusian farmers. Time devoted to the farm and education level were the most important explanatory variables.

Key words: Logit model, Dummy variable, Farm income.

Introducción

En el diseño de políticas agrarias es de gran utilidad la separación de agricultores que sólo tienen como fuente de ingreso la actividad agraria de aquellos que disponen de una fuente de ingresos adicional. Con esta información se podrían determinar diferencias de rentas entre exclusivamente agricultores y otros sectores de la sociedad. Desde el punto de vista de uno de los objetivos de la Política Agraria Común, el incremento de rentas de los agricultores, el presente estudio añade elementos de juicio para evaluar qué tipo de agricultores (tama-

ño, orientación técnica productiva, edad, etc.) son en mayor medida beneficiados o perjudicados por iniciativas concretas.

Para la caracterización de estos agricultores se utilizó un modelo econométrico cuya variable dependiente es dicotómica. Dicha variable toma el valor uno cuando el porcentaje de la renta de los agricultores que proviene de la explotación agraria es el cien por cien, y cero cuando el agricultor tiene otras fuentes de ingreso. Un modelo econométrico logit (ALDRICH y NELSON, 1984) es elegido para explicar la relación de la variable renta con otras variables de tipo socioeconómico. Puede encontrarse este tipo de modelo en trabajos previos como DAOULI y DEMOUSSIS (1995), MILLÁN y RUIZ (1987) y GARCÍA *et al.* (1983).

Los datos

Durante el verano de 1996 fue llevada a cabo una encuesta en toda Andalucía. El procedimiento para la selección de los agricultores fue el de muestreo estratificado. Cuando la población presenta una gran heterogeneidad, como es el caso de la agricultura andaluza, dicho procedimiento es mucho más eficiente que el muestreo aleatorio simple (WONNACOTT, 1990). Los criterios de distribución del número de encuestas en cada comarca de las 55 comarcas andaluzas fueron tres: valor añadido del sector agrario (no ganadero), número de explotaciones agrarias y diversidad de cultivos. Para construir el modelo se utilizaron 531 observaciones.

Las variables que se han considerado en el estudio aparecen en el cuadro 1.

Estas variables recogen información socioeconómica del agricultor y el tipo de explotación. Con respecto a las tres variables sociológicas, la variable edad intenta mostrar si los agricultores de diferentes edades tienen otras fuentes de renta además de la propia actividad agraria; el nivel de educación, parece, a priori, positivamente

CUADRO 1
VARIABLES DEL MODELO LOGIT
TABLE 1
VARIABLES OF THE LOGIT MODEL

Variable	Tipo de variable	Descripción
Renta	Variable dependiente, dicotómica	1= 100% renta viene de su explotación agraria, 0= hay ingresos ajenos a la explotación agraria
Tiempo	Variable explicativa, limitada	Porcentaje de su actividad laboral total anual que dedica a la explotación agraria
Herbaceos	Variable explicativa, ficticia	1= explotación exclusivamente de herbáceos, 0= no herbáceos en la explotación
Cítricos	Variable explicativa, ficticia	1= presencia de cítricos, 0= no cítricos
Olivar	Variable explicativa, ficticia	1= presencia de olivar, 0= no olivar
Edad	Variable explicativa, continua	Edad del agricultor
Educación	Variable explicativa, categórica	Nivel educativo: 1= sin estudios hasta 5= universitarios
Superficie	Variable explicativa, continua	Superficie de la explotación en hectáreas
Propiedad	Variable explicativa, limitada	Porcentaje de la explotación en propiedad
Pastos	Variable explicativa, limitada	Porcentaje de pastos en la explotación
Irrigación	Variable explicativa, limitada	Porcentaje de regadío en la explotación

relacionada con el porcentaje de renta que proviene de actividades agrarias; por último, el porcentaje de tiempo dedicado a la explotación debería estar, de igual forma, positivamente relacionado con la variable renta.

Las restantes variables están relacionadas con el tipo de explotación o actividad desarrollada. El tamaño de la explotación podría tener una relación positiva con la variable renta considerando que una mayor explotación proporciona mayores ingresos y una pequeña haría buscar al agricultor fuentes alternativas de renta. El porcentaje de tierra en propiedad podría moverse en la misma dirección ya que puede haber agricultores que arriendan tierras y reciben toda su renta por esta vía.

En cuanto a las actividades llevadas a cabo en la explotación, herbáceos, críticos, olivar o pastos, sólo en el caso de esta última podría esperarse un claro efecto negativo en la variable renta, ya que estos producen un menor ingreso, al contrario que el porcentaje de regadío en la explotación.

Los datos provienen de una encuesta realizada en el verano de 1996, en toda Andalucía, a 531 agricultores, dentro del proyecto INIA SC94-095.

Fundamentación teórica

En estudios cuyos datos provienen de encuestas se encuentra con frecuencia que la variable que se desea explicar, y_i, es de naturaleza cualitativa o varía dentro de un rango limitado. La mayoría de los modelos econométricos (MADDALA, 1983) que tratan este tipo de variables pertenecen a una de las siguientes categorías: (a) *modelos de variable truncada*, donde la muestra sólo cubre una parte de la población y, por tanto, sólo podemos observar y_i cuando $y_i \leq c$ (ó $y_i \geq c$) siendo c una constante determinada; (b) *modelos de variable censada*, en este caso se dispone de los valores de las variables explicativas para $y_i \leq c$; (c) *modelos de variable dicotómica*, donde la variable dependiente es binaria, tomando valores cero o uno. El modelo desarrollado en este artículo pertenece a la última categoría.

Una de las posibles aproximaciones a modelos econométricos de variable dependiente binaria es el modelo lineal probabilístico (GREENE, 1993). Considerando por ejemplo que y_i toma el valor uno cuando el individuo posee una casa y cero en cualquier otra circunstancia, el vector x_i representa el conjunto de variables explicativas (renta familiar, edad, número de hijos, etc.), β: es el vector de coeficientes a estimar y u_i la disturbancia, el modelo lineal probabilístico tiene la forma:

$$y_i = \beta' x_i + u_i$$

Este modelo puede considerarse como la probabilidad de poseer una casa, $y_i = 1$, para unos valores determinados de x_i, esto es, la probabilidad condicional, $P(y_i=1|x_i)$:[1]. La aplicación directa de mínimos cuadrados ordinarios (MCO) presenta algunas dificultades, violando algunos de los supuestos en que se basa dicho procedimiento (CRAMER, 1991).

Funciones de distribución acumulativas

El principal problema con el modelo lineal probabilístico es que la probabilidad crece

1. Para facilitar la notación utilizaremos en adelante P_i en lugar de P_i $(y_i=1|X_i)$.

linealmente con x_i. Este comportamiento no es muy realista ya que a partir de un cierto valor de x_i la probabilidad P_i no debería de aumentar significativamente. Así, el modelo debe cumplir dos requisitos: a) los valores estimados deben estar dentro del intervalo (0,1), y b) la relación que liga x_i y P_i no sea lineal. El modelo logit tiene estas dos características, matemáticamente:

$$P_i = E(y_i=1|x_i) = \frac{1}{1+e^{\beta' xi}}$$

Estimador de máxima verosimilitud

Los parámetros del modelo logit suelen estimarse mediante el estimador de máxima verosimilitud (EMV), en lugar de MCO[2]. Considerando $Pi=P(y_i=1|x_i)$, entonces $P(y_i=0|x_i)= 1- Pi$, y la probabilidad de obtener y_i, sea 0 ó 1, viene dada por $P(y_i|x_i)= P_i^{yi}(1-P_i)^{1-yi}$. La función logarítmica a maximizar puede expresarse como sigue:

$$\text{Log } L(y|x,\beta) = \sum_{i=1}^{N} \left[yi \log Pi + (1-yi)\log(1-P_i) \right]$$

La condición necesaria para maximizar esta función es:

$$\frac{\partial \log L\,(y|x,B)}{\partial \beta} = 0$$

Ya que esta ecuación no es lineal es necesaria la utilización de un algoritmo iterativo para obtener el valor de los parámetros. EMV presenta propiedades similares a las de MCO, ambos estimadores son insesgados y eficientes y generan residuos que se distribuyen como una normal. La principal diferencia es que EMV es un estimador no lineal de propiedades asintóticas.

Resultados y discusión

Regresión logística

A continuación, el cuadro 2 muestra el valor del estadístico t para las variables que no fueron rechazadas por ser estadística-

CUADRO 2
REGRESIÓN LOGÍSTICA
TABLE 2
LOGISTIC REGRESSION

Variable	Coeficientes	t-ratios	Probabilidad
C	-15,10363	-5,282899	0,0000
TIEMPO	0,172705	6,131979	0,0000
EDAD	-0,032673	2,104572	0,0358
EDUCACIÓN	0,481115	2,623653	0,0090
SUPERFICIE	0,003977	2,062417	0,0397
PASTOS	-0,026813	-2,885076	0,0041

Logaritmo de verosimilitud: -133,6789

2. Mediante MCO obtenemos la misma estimación del los parámetros que con EM cuando los residuos se distribuyen como una normal. En el caso de un modelo logístico, MCO puede ser aplicado pero el estimador no es líneal y además insuficiente.

mente significativas para un nivel del 5 por ciento.

Para comprobar si un conjunto de coeficientes son significativos es necesario utilizar la submatriz de covarianzas de los parámetros a comprobar. Para este propósito se utilizó el test de Wald, cuyo estadístico se calcula mediante la siguiente expresión:

$$w = \hat{\beta}'_* V_*^{-1} \hat{\beta}_*$$

Teniendo en cuenta el valor del estadístico t en el cuadro 2 podemos seleccionar los coeficientes no significativos y junto con la submatriz de covarianzas de estos coeficientes, obtenemos el estadístico de Wald, w, para $\hat{\beta}_*$ = (-0,0924, -0,3820, 0,7861, -0,0093, 0,0001), w= 6,10.

El estadístico de Wald se distribuye como una chi cuadrado con L (número de parámetros a comprobar) grados de libertad. $\chi^2_{5,0.05} = 11,10 > 6,10$. No rechazamos la hipótesis nula de todos los coeficientes igual a cero.

Así, la probabilidad de encontrar un agricultor cuya renta provenga exclusivamente de la actividad agraria, renta = 1, dado un conjunto de valores de las variables explicativas, se define mediante la siguiente función de probabilidad:

$$Pi = P(renta = 1|b'x) = \frac{1}{1+e^{-(-15.104+0.173*tiempo-0.033*edad+0.481*edu+0.004sup-0.027*past)}}$$

Teniendo en cuenta el exponente negativo de e, las variables *tiempo* y *educación* tienen el mayor peso en la variable renta. También influye positivamente la variable *superficie*, aunque en menor medida. Por el contrario, *edad* y *pastos* tienen un efecto negativo en la probabilidad de que toda la renta del agricultor venga de la explotación agraria.

Precisión de la regresión

La prueba F es tradicionalmente usada para comprobar si la hipótesis nula de que todos los coeficientes son cero es rechazada o no. En el modelo logit se dispone de una prueba similar (DAVIDSON y MACKINSON, 1984; ENGLE, 1984) que utiliza la ratio definida como: c= $-2(\log L_0 - \log L_1)$. Donde L_1 es el valor de la función de probabilidad en el modelo logit con todas las variables y L_0 el correspondiente valor en el modelo reducido, esto es, todos los coeficientes cero, excepto el término independiente. El resultado habitual de los paquetes econométricos es el logaritmo de este valor, como en este caso, al pie del cuadro 2 tenemos el logaritmo de L_1 (Log likelihood). Este ratio se distribuye aproximadamente como una chi cuadrado con K-1 grados de libertad (siendo K el número de parámetros a comprobar) cuando la hipótesis nula es verdadera.

No es necesario repetir la regresión sólo con la constante para obtener $\log L_0$, ya que podemos utilizar la siguiente igualdad:

$$\log L_0 = N_0 * \log \frac{N_0}{N} + N_1 * \log \frac{N_1}{N}$$

con N representado tamaño de la muestra y N_0 y N_1 número de observaciones pra renta=0 y renta= 1, respectivamente. En este caso,

$$\log L_0 = 304 * \log \frac{304}{531} + 227 * \log \frac{227}{531}$$

$$= -362,46$$

por tanto

$$c = -2[-362,46-(-133,68)] = 457,56$$

Como $\chi^2_{5,0.05} = 11{,}10$, rechazamos la hipótesis nula de que todos los coeficientes son cero.

Otro dato usado generalmente para medir la bondad del ajuste es el valor de R^2, la ratio varianza explicada por el modelo / varianza total. Aunque en el modelo logit no hay un estadístico similar, sí existen otros medios para obtener esta información. Para este propósito es necesario utilizar los valores L_1 y L_0, a partir de sus logaritmos. Uno de ellos se define como $R^2=1-(L_0/L_1)^{2/n}$. Utilizando los logaritmos: $\log L_1 = -133{,}68 \rightarrow L_1 = 8{,}78*10^{-59}$ y $\log L_0 = -362{,}40 \rightarrow L_0 = 4{,}09*10^{-158}$. Así, tenemos: $R^2=1-(4{,}09*10^{-158}/8{,}78*10^{-59})^{2/531}= 0{,}58$.

Aunque el valor de R^2 no parezca a primera vista muy elevado hay que tener en cuenta que en este tipo de modelos el límite superior no es uno. Calculando este límite,

$0 \leq R^2 \leq 1 - L_0^{2/n}$,

$0 \leq R^2 \leq 1 - 0{,}26 = 0{,}74$

Otra medida es el llamado pseudo-R^2 definido por CRAGG y UHLER (1970),

pseudo-$R^2 = (L_1^{2/n} - L_0^{2/n})/(1 - L_0^{2/n})$,

pseudo-$R^2 = 0{,}60$

O la formulación de pseudo-R^2 propuesta por MCFADDEN (1974),

pseudo-$R^2 = 1 - \log L_1/\log L_0$,

pseudo-$R^2 = 1 - 0.37 = 0{,}63$.

Todas estas medidas de precisión de la regresión no son universalmente aceptadas aunque las tres coinciden en mostrar que el modelo propuesto explica gran parte de la varianza total. Un método alternativo para medir la capacidad predictiva del modelo consiste en atribuir el valor 1 a la variable renta cuando la probabilidad estimada, P_i, sea mayor que 0,5 y el valor 0 cuando sea menor. Así, para cada agricultor calculamos la probabilidad de que perciba toda la renta de la actividad agraria y lo comparamos con el valor observado de renta, el siguiente cuadro muestra los resultados para todos los agricultores:

De acuerdo con este cuadro, de los 304 agricultores que no perciben toda su renta de la actividad agraria, el modelo predice 252. Matemáticamente:

$$P(\hat{P}_i < 0{,}5 | yi = 0) = \frac{258}{304} = 0{,}85$$

$$P(\hat{P}_i > 0{,}5 | yi = 1) = \frac{221}{227} = 0{,}97$$

CUADRO 3
VALORES DE RENTA ESTIMADOS FRENTE A OBSERVADOS
TABLE 3
ESTIMATED AND OBSERVED DATA

		Valores estimados		Total
		0	1	
Valores	0	252	52	304
observados	1	6	221	227
		258	273	531

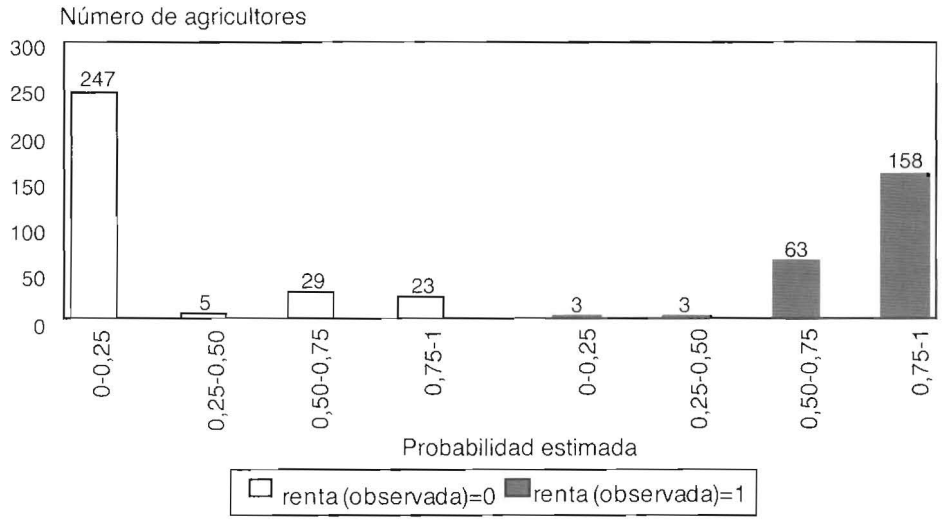


Figura 1. Distribución de las probabilidades estimadas mediante el modelo logit
Figure 1. Probability distribution of the estimated values

Así, la capacidad predictiva total del modelo puede calcularse como: (252+221)/531= =0,89. Otro elemento a tener en cuenta es que la mayoría de las probabilidades estimadas se encuentra próxima al valor observado de la variable renta; este punto es mostrado en la siguiente distribución de probabilidades en la figura 1.

Puede observarse como para renta observada=0 encontramos 247 estimaciones inferiores a 0,25. Las estimaciones incorrectas para renta=0 son 29+23, ya que la probabilidad estimada es superior a 0,5.

Sensibilidad del modelo

Es posible determinar el efecto en la probabilidad Pi, P(renta=1), de las variables explicativas. Por ejemplo, la siguiente figura muestra el cambio en dicha probabilidad con respecto a la variable tiempo y nivel educativo. Para el resto de variables explicativas en el modelo, edad, superficie y pasto, se utilizan sus valores medios. Así, en el modelo,

$$Pi = P(renta = 1|b'x) = \frac{1}{1+e^{-(-15,104+0,173*tiempo-0,033*edad+0,481*edu+0,004*sup-0,027*past)}}$$

utilizamos los valores medios edad= 49,8, superficie= 82,3 ha y porcentaje de pastos= 4,7. La variable tiempo tomó los valores 0-100 y nivel educativo 1-5. La figura 2 recoge esta información,

La figura muestra como para un valor de tiempo inferior al 65 por ciento, la probabilidad de encontrar un agricultor que tenga sólo ingresos agrarios es cero. Para un valor de tiempo igual al 85 por ciento dicha probabilidad varía desde el 0,2 para un nivel educativo mínimo, hasta 0,6 cuando

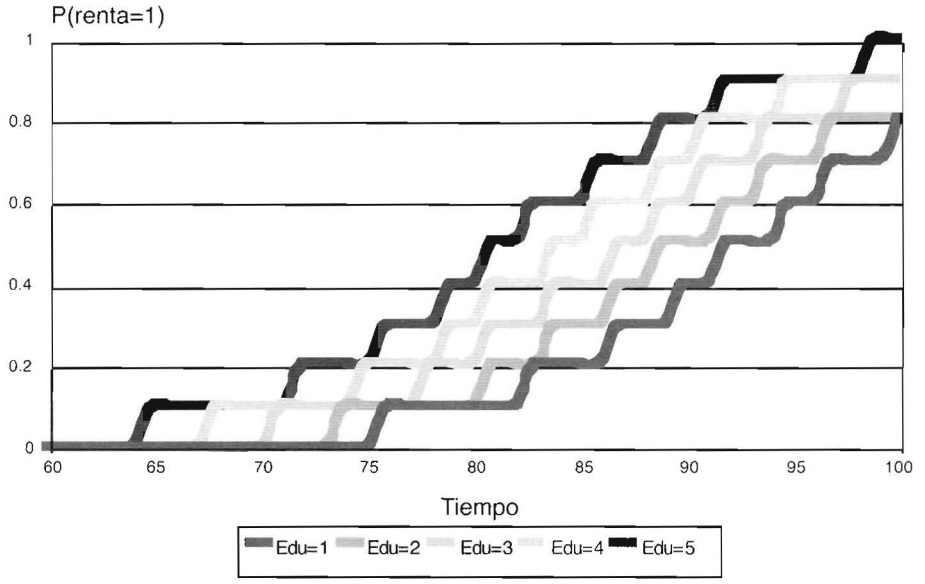


Figura 2. Cambio en la probabilidad estimada de renta exclusiva agraria para las variables explicativas Tiempo y Nivel educativo

Figure 2. Change in estimated probability for agrarian rent according to the explaining variables Time and Education level

el nivel educativo es máximo. Este análisis puede repetirse para otras variables (valor medio de nivel educativo= 1,9 y tiempo= 68,7), las variables continuas pueden convertirse en categóricas para analizarlas como el nivel educativo.

Conclusiones

El presente trabajo propone un modelo econométrico que explica la probabilidad de encontrar un agricultor cuya renta provenga exclusivamente de la actividad agraria, dados unos determinados valores de las variables explicativas *tiempo* (porcentaje de la actividad laboral del agricultor que dedica a la explotación), *edad*, *nivel educativo*, *superficie* y *pastos* (porcentaje de la explotación cubierta de pastos). De estas variables, *nivel educativo* y *tiempo* son, con diferencia, las variables que mayor peso tienen en el incremento de dicha probabilidad. Las variables *edad* y *pastos* tienen, por el contrario, un efecto negativo en esta probabilidad, aunque de menor importancia.

Son de destacar las variables que no parecen tener relación, contra lo que a priori podría esperarse, con la variable dependiente (renta exclusivamente agraria). Estas fueron: *regadío* (porcentaje de la explotación agraria en regadío), *propiedad* (porcentaje en propiedad) y orientación técnico productiva de la explotación: *herbáceos* (explotación exclusivamente con cultivos herbáceos), *cítricos* (presencia de cítricos)

y *olivar* (presencia de olivar).

Agradecimiento

Los autores quieren expresar su agradecimiento a los revisores anónimos por sus comentarios y sugerencias.

Bibliografía

ALDRICH J., y NELSON F., 1984. Linear Probability, Logit and Probit Models. Sage Publications, Beverly Hills. California.

CRAGG J.G., UHLER R., 1970. The demand for automobiles. Canadian Journal of Agricultural Economics, 3, 386-406.

CRAMER J.S., 1991. The Logit Model for Economists. Edward Arnold Publishers. London.

DAOULI J., DEMOUSSIS M., 1995. Trabajo fuera de las explotaciones en la agricultura griega: una aplicación empírica. Investigación Agraria: Economía, 10 (1), 77-89.

DAVIDSON R., MACKINNON, 1984. Convenient Specification Tests for Logit and Probit Models. Journal of Econometrics, 25, 241-262.

ENGLE R., 1984. Wald, likelihood Ratio, and Lagrange Multiplier Test in Econometrics. En: Handbook of Econometrics. Vol. 2, Z. Griliches y M. Intriligator (Ed.). Ed. North Holland, Amsterdam.

GARCÍA P., SONKA S. T., MAZZACO M. A., 1983. A Multivariate Logit Analysis of Farmers' Use of Finantial Information. American Journal of Agricultural Economics, 65, 136-141.

GREENE W.H., 1993. Econometric Analysis. Prentice-Hall International. New Jersey.

MADDALA G.S., 1983. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University press. New York.

McFADDEN D., 1974. Conditional Logit Analysis of Qualitative Choice Behaviour. En: Frontiers in econometrics, 105-142. P. Zarembka (Ed.). Ed. Academic Press. New York.

MILLAN J. A., RUIZ P., 1987. Modelos Logit de Adopción de Innovaciones en Invernaderos de Almería. Investigación Agraria: Economía, 2, 115-125.

WONNACOTT T.H., y WONNACOTT R.J., 1990. Introductory Statistics for Business and Economics. John Wiley. Singapore.

(Aceptado para publicación el 29 de septiembre de 1998)