

# APLICACION DE MODELOS DE ELECCION DISCRETA PARA LA DETECCION DEL FRAUDE EN LA IMPOSICION SOBRE LA RENTA (\*)

José Luis RAYMOND BARA  
Teófilo VALDES SANCHEZ

## I. CONSIDERACIONES PREVIAS

**L**A imposición sobre la renta constituye uno de los pilares básicos de los sistemas fiscales modernos. No obstante, la gestión de un impuesto de tal naturaleza es compleja, en la medida en que puede resultar difícil para la Administración disponer de conocimiento detallado de las diversas fuentes de renta y de las circunstancias personales que afectan a los distintos contribuyentes individuales. Para que el impuesto cumpla con los requisitos de equidad que la lógica del sistema exige, es preciso disponer de elevada capacidad inspectora. Sucede, sin embargo, que las posibilidades de la Administración son siempre limitadas.

Por ello, en diversos países, como Francia y Estados Unidos, se ha ensayado el empleo de técnicas estadísticas que permitan una primera clasificación de contribuyentes entre «probables» defraudadores y no defraudadores. Obtenida esta clasificación, la actividad inspectora debe centrarse, principalmente, en los primeros. En definitiva, en ausencia de otra información, y dado que por falta de medios no es posible llevar a cabo una investigación exhaustiva de la totali-

dad de contribuyentes, se trata de ofrecer un criterio a la Administración que permita clasificarlos en los dos grupos mencionados.

Con relación a España, un estudio de Valdés *et al.* (1982) llevó a cabo la aplicación del «Análisis Discriminante» para clasificar las declaraciones de renta entre fraudulentas y no fraudulentas. El punto de partida fue la extracción de una muestra completamente aleatoria de contribuyentes que fueron objeto de inspección detallada. Se comprobó quiénes eran defraudadores y quiénes no lo eran, y, por medio del análisis estadístico de las características personales de tales contribuyentes reflejadas en la propia declaración, se realizó la estimación de la correspondiente función discriminante.

En este trabajo se ha creído conveniente la ampliación del análisis, ensayando un tipo de modelos que progresivamente van siendo utilizados con mayor profusión en el campo de la econometría. Son los modelos «Probit» y «Tobit».

El punto de partida es una muestra aleatoria de 1.270 contribuyentes correspondientes a la declaración ordinaria de la imposición sobre la renta del ejercicio 1979. Estos contribuyentes fueron objeto de una inspección

detallada. Definiendo el fraude como la aparición de una diferencia en exceso a 20.000 pesetas en la cuota, se obtuvo el resultado de que un 27 por 100 de declaraciones eran fraudulentas. El umbral de 20.000 pesetas de diferencia en la cuota para establecer la separación entre declaraciones fraudulentas y no fraudulentas obedece a los costes estimados de inspección.

Utilizando esta información, se trata de estimar un modelo (Modelo «Probit») que cuantifique la probabilidad de que una declaración sea fraudulenta atendiendo a ciertos epígrafes de la declaración rellenos o no rellenos por el contribuyente. A un segundo nivel, se trata de estimar un modelo (Modelo «Tobit») que no sólo cuantifique la probabilidad de que una declaración sea fraudulenta, sino que también permita llevar a cabo una evaluación aproximada del fraude.

Una vez formulados y estimados los modelos utilizando esta muestra de 1.270 observaciones, se dispone de una muestra adicional de 978 observaciones, que no ha sido utilizada en la fase de selección y estimación de los modelos, y que se emplea únicamente para validar los modelos previamente estimados.

La finalidad del trabajo es que, después de aceptar un modelo de comportamiento del contribuyente, la misma formulación pueda ser aplicada a las declaraciones de renta de otros ejercicios económicos, de manera que los modelos «Probit» y «Tobit» permitan llevar a cabo una ordenación de declaraciones atendiendo a la probabilidad de fraude y a su cuantía estimada. En este sentido, la capacidad inspectora de la Administración fija el número total de declaraciones

que pueden ser objeto de comprobación y, en ausencia de otra información *a priori*, la selección de declaraciones a inspeccionar no se realiza de forma puramente aleatoria, sino de acuerdo con la probabilidad de fraude que el modelo predice. El planteamiento puede ser útil para el caso de grandes ciudades, como Madrid o Barcelona, en que, en general, el grado de conocimiento que la Administración tiene de los contribuyentes individuales es muy limitado o inexistente y su selección es llevada a cabo por procedimientos basados en la experiencia. En estas circunstancias, disponer de un modelo de tal naturaleza contribuye a mejorar los resultados que puedan derivarse de una selección de contribuyentes puramente aleatoria o heurística.

Otra ventaja de disponer de un modelo de selección de contribuyentes es que el proceso en su totalidad puede ser informatizado. Ello significa que, sin apenas destinar recursos humanos, el conjunto de declaraciones de la imposición sobre la renta de un determinado ejercicio puede ser ordenado de mayor a menor probabilidad de fraude. Sobre tales declaraciones ordenadas puede centrarse la inspección.

Un problema que se plantea es el de corregir los posibles efectos que la inflación pueda tener sobre el modelo de selección de contribuyentes. En efecto, entre 1979 y un ejercicio posterior las bases fiscales deben experimentar una apreciable variación como consecuencia del crecimiento de los precios. A este respecto, el procedimiento que generalmente se sigue es el de reconvertir todas las magnitudes a unidades monetarias del año base que se utiliza a efectos de la estimación del modelo, empleando como de-

flactor el crecimiento monetario del epígrafe considerado. A título ilustrativo, si los rendimientos netos del trabajo entre dos fechas han aumentado en un 50 por 100, el deflactor correspondiente a aplicar a los rendimientos del trabajo del segundo período, a efectos de transformarlos en rendimientos del trabajo del año base, sería de 1,5. La experiencia al respecto es que tal metodología ofrece resultados válidos, y que los modelos estimados, después de esta corrección, no experimentan, en general, cambios estructurales significativos. En efecto, cuando después de transcurridos algunos años el modelo es objeto de reestimación, los resultados de otros países sugieren que la ordenación de declaraciones de acuerdo con el nuevo modelo sigue siendo similar a la que se obtenía con el modelo inicial, después de establecer la corrección de deflactar las rentas para convertirlas a unidades monetarias del año base. Si bien los resultados de estabilidad del modelo a los que se hace referencia corresponden al empleo del «Análisis Discriminante», no hay razón objetiva para suponer que esta consideración no sea extensible a los modelos «Probit» y «Tobit». Parece, pues, que únicamente un cambio estructural importante en el sistema fiscal (bien sea de naturaleza legal o de otro tipo) puede convertir al modelo en inoperante. De cualquier forma, los modelos deben ser reestimados cada cierto tiempo, que en los países donde se vienen utilizando varía entre cinco y diez años.

En la exposición que sigue se efectúa, en primer lugar, una presentación de los modelos «Probit» y «Tobit», centrando la exposición de tales modelos en el

análisis de la problemática objeto de estudio. Posteriormente, se ofrecen los resultados de su estimación para las declaraciones de renta en España del ejercicio de 1979 utilizando la muestra a la que ya se ha hecho referencia. No obstante, el modelo finalmente estimado no puede exponerse de forma detallada dado que, en tal caso, en la medida en que fuese conocido por el contribuyente, dejaría de ser eficaz para detectar el fraude. Por último, el trabajo concluye con unas conclusiones provisionales.

Interesa destacar que este artículo ha sido escrito antes de disponer de resultados definitivos, por lo que sus conclusiones deben necesariamente tener un carácter provisional.

## II. FORMULACION DE LOS MODELOS «PROBIT» Y «TOBIT»

Una presentación general de los modelos denominados «Modelos de Elección Discreta» (o también «Modelos con Funciones de Respuestas Categóricas»), término que engloba como casos especiales a los modelos «Probit» y «Tobit», puede hallarse en diversos *survey*, como el de Amemiya (1981) o el reciente libro de Maddala (1983). La exposición que seguidamente se efectúa va a centrarse sólo en los aspectos conceptuales y poniendo el énfasis en aquellos extremos que interesan a efectos de este trabajo.

### 1) Formulación del modelo «Probit»

El modelo «Probit» aplicado en este contexto trata de cuan-

tificar la probabilidad de que un determinado contribuyente sea defraudador. Existen diversas alternativas para justificar el empleo de este modelo. No obstante, la teoría de la utilidad, según la sugerencia de Mc. Fadden (1973, 1976), puede resultar adecuada a efectos expositivos. Adaptando la formulación general al caso en que estamos interesados, podemos considerar que el contribuyente  $i$  se enfrenta ante dos opciones: La opción 1 es la de defraudar y la opción 0 es la de no defraudar. Se supone que el contribuyente eligirá una de estas dos opciones atendiendo a las respectivas utilidades netas esperadas.

Así, considerando una función de utilidad lineal en los parámetros, la utilidad que el contribuyente  $i$  deriva de la opción de defraudar puede representarse a través de:

$$U_{i1} = X'_{i1} \beta_1 + e_{i1}$$

En este caso,  $U_{i1}$  es la utilidad que el contribuyente  $i$  espera que le reporte la opción de defraudar,  $X'_{i1}$  es un vector de características del contribuyente  $i$  relevantes a efectos de definir las ventajas esperadas del fraude (a título ilustrativo, éste podría ser el caso del nivel de renta: cuanto más elevado es el nivel de renta, y dado un tipo impositivo marginal positivo y creciente, mayor cabe esperar que sea el incentivo a defraudar impuestos),  $\beta_1$  es un vector de parámetros poblacionales asociados a estas características, y por último,  $e_{i1}$  es una variable aleatoria para la que se supone normalidad, y que interviene de forma aditiva comportando que la función de utilidad sea estocástica. La justificación de introducir un elemento de perturbación aleatoria en la función de utilidad

generalmente discurre por dos vías: En primer lugar, como en el modelo de regresión *standard*, la perturbación aleatoria puede recoger la influencia de las variables omitidas o no observables. A título ilustrativo, el distinto grado de sentido cívico de los contribuyentes podría quedar recogido por la perturbación aleatoria. En efecto, dado que los parámetros poblacionales  $\beta$  se suponen constantes entre los distintos contribuyentes, y dado que en igualdad de condiciones no todos ellos son igualmente proclives al fraude, la distinta escala de valores de los contribu-

yentes individuales podría quedar reflejada por la perturbación aleatoria. En segundo lugar, es posible conceptuar el comportamiento del individuo como inherentemente estocástico.

Al igual que sucede en el modelo de regresión, es posible que los elementos  $X'_{i1}$  sean logaritmos, productos, cocientes o cualquier otra combinación de las características del individuo  $i$  asociados a la opción 1.

La utilidad asociada a la opción 0 vendrá dada por:

$$U_{i0} = X'_{i0} \beta_0 + e_{i0}$$

CUADRO N.º 1

DETECCION DEL FRAUDE EN UNIDADES MONETARIAS

RESULTADOS DENTRO DE LA MUESTRA					
PROBABILIDAD SELECCIONADA	% de la muestra inspeccionada (1)	% de fraude detectado (2)	Ratio (2) (1)	% de aciertos (3)	Ratio (3) 27,46
P > 0,9	0,87	6,98	8,03	90,91	3,31
P > 0,8	1,26	8,60	6,82	81,25	2,95
P > 0,7	2,36	10,78	4,56	66,67	2,42
P > 0,6	3,78	14,95	3,95	68,75	2,50
P > 0,5	5,51	19,34	3,50	68,57	2,49
P > 0,4	8,89	27,80	3,12	64,60	2,35
P > 0,3	14,95	40,26	2,69	59,47	2,16
P > 0,2	32,49	63,11	1,94	49,64	1,80
P > 0,1	93,15	97,96	1,05	28,80	1,04
P > 0,0	100,00	100,00	1,00	27,46	1,00

RESULTADOS FUERA DE LA MUESTRA					
PROBABILIDAD SELECCIONADA	% de la muestra inspeccionada (1')	% de fraude detectado (2')	Ratio (2') (1')	% de aciertos (3')	Ratio (3') 23,93
P > 0,9	0,61	5,36	8,78	83,33	3,48
P > 0,8	0,92	7,43	8,07	77,78	3,25
P > 0,7	1,43	7,86	5,49	64,29	2,68
P > 0,6	2,56	11,70	4,57	68,00	2,84
P > 0,5	3,58	15,08	4,21	65,71	2,74
P > 0,4	5,62	20,28	3,60	61,82	2,58
P > 0,3	9,61	28,50	2,96	56,38	2,35
P > 0,2	21,57	49,82	2,30	49,29	2,05
P > 0,1	91,10	96,97	1,06	25,36	1,05
P > 0,0	100,00	100,00	1,00	23,93	1,00

En este contexto, definiendo la variable no observable  $Y^*$  como la diferencia de utilidades de las dos opciones, se tiene:

$$\begin{aligned} Y_i^* &= U_{i1} - U_{i0} = \\ &= X'_{i1} \beta_1 - X'_{i1} \beta_0 + e_{i1} - e_{i0} = \\ &= Z'_i \alpha + \varepsilon_i \end{aligned}$$

en donde  $Z'_i \alpha$  recoge el componente sistemático de las respectivas funciones de utilidad y  $\varepsilon_i$  es una nueva variable aleatoria cuya distribución sigue siendo normal.

En el caso del modelo «Probit» se observa la variable aleatoria discreta  $Y_i$  que adopta los valores «cero» si el individuo no defrauda y «unitario» si el individuo defrauda. Es decir:

$$\begin{aligned} Y_i &= 0 && \text{Si } U_{i1} < U_{i0} \\ Y_i &= 1 && \text{Si } U_{i1} > U_{i0} \end{aligned}$$

Ello significa que si la utilidad esperada de la opción de defraudar es inferior a la utilidad esperada de la opción de no defraudar (la utilidad puede definirse en términos suficientemente amplios para dar cabida a todos aquellos factores distintos de los estrictamente económicos), el individuo no defrauda y se observa  $Y_i = 0$ . A diferencia, si la utilidad esperada de la opción de defraudar excede a la de la opción de no defraudar, el individuo defrauda y se observa  $Y_i = 1$ .

Partiendo de este enfoque, la probabilidad de que el individuo  $i$  no sea defraudador viene dada por:

$$\begin{aligned} \text{Prob. } (Y_i = 0) &= \text{Prob. } (U_{i0} > U_{i1}) = \\ &= \text{Prob. } (Y_i^* < 0) = \\ &= \text{Prob. } (\varepsilon_i < -Z'_i \alpha) = \\ &= \text{Prob. } \left( \frac{\varepsilon_i}{\sigma} < -Z'_i \alpha^* \right) = \\ &= \Phi (-Z'_i \alpha^*) = 1 - \Phi (Z'_i \alpha^*) \end{aligned}$$

en donde  $\sigma$  es la desviación *standard* del elemento de per-

turbación aleatoria  $\varepsilon_i$ ,  $\alpha^*$  se define como el cociente entre  $\alpha$  y  $\sigma$ , y  $\Phi (Z'_i \alpha^*)$  es el valor de la función de distribución asociada a una normal (0,1).

Por diferencia, la probabilidad de que el individuo  $i$  sea defraudador vendrá representada por:

$$\text{Prob. } (Y_i = 1) = \Phi (Z'_i \alpha^*)$$

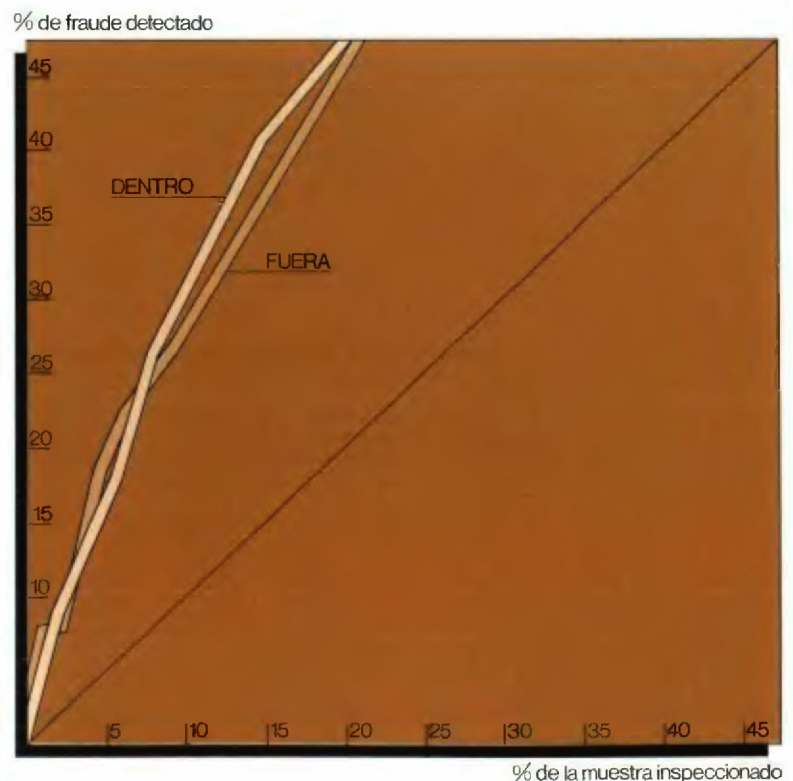
Suponiendo que se dispone de una muestra observada de  $N$  individuos, de los que se conocen sus características personales así como si efectivamente han sido defraudadores o no, es posible formar la función de verosimilitud con objeto de llevar

a cabo la estimación de los parámetros del modelo.

## 2) Formulación del modelo «Tobit»

La formulación «Tobit» permite determinar, no sólo la probabilidad de fraude del contribuyente  $i$ , sino también el valor del fraude predicho para tal contribuyente. En este caso la variable observada adopta valores «cero» si el contribuyente no defrauda y valores continuos (es decir, el importe del fraude) si el contribuyente es defraudador.

GRAFICO 1  
DETECCION DEL FRAUDE  
EN UNIDADES MONETARIAS



En consecuencia, la variable observable  $Y_i$  se define de la siguiente forma:

$$Y_i = \begin{cases} F_i \text{ (Fraude observado)} \\ \text{Si: } F_i = Z_i' \alpha + \varepsilon_i > 0 \quad [1] \\ 0 \text{ (No fraude)} \\ \text{Si: } Z_i' \alpha + \varepsilon_i < 0 \quad [2] \end{cases}$$

Así, tal modelo utiliza más información que el precedente, dado que la misma hace referencia a la cantidad defraudada en el supuesto de que el individuo  $i$  sea efectivamente defraudador. No obstante, si el individuo no defrauda, se observa un fraude nulo. La característica del modelo, acorde con la realidad, es que el fraude puede ser cero, pero nunca negativo, y la principal ventaja radica en que, como antes ya se ha señalado, un modelo de tal naturaleza puede predecir la probabilidad de fraude a través de la relación [2], y si el individuo es defraudador, posibilita llevar a cabo una evaluación de la cantidad defraudada por medio de la ecuación continua [1].

En definitiva, y tal como reflejan las ecuaciones [1] y [2],

se supone que el proceso de generación de los datos obedece a dos regímenes según que el individuo sea defraudador, en cuyo caso es relevante la regresión continua, o no defraudador, en que la relación  $Z_i' \alpha + \varepsilon_i < 0$  sólo permite determinar la probabilidad de fraude.

Como en el caso de la modelización «Probit», es posible formar la función de verosimilitud y llevar a cabo la estimación de los parámetros poblacionales del modelo maximizando dicha función.

### III. RESULTADOS DE LA ESTIMACION

En la actualidad se está utilizando la evaluación de los modelos «Probit» y «Tobit». Únicamente se dispone de resultados definitivos para las aproximaciones a los respectivos modelos a través de la estimación del «Modelo Logit» y de una aproximación por Mínimos Cuadrados al «Modelo Tobit». Ello significa que la capacidad de discrimina-

ción del modelo que seguidamente se expone es más pobre de la que cabe esperar a partir de la aplicación de otras técnicas de estimación más sofisticadas. El motivo es doble:

En primer lugar, como antes ya se ha señalado, el método de estimación utilizado no es el más adecuado, lo que repercute en detrimento de la bondad del modelo.

En segundo lugar, la selección de variables no es definitiva, dado que la significatividad de las mismas puede resultar alterada al pasar a otros métodos de estimación, lo que puede hacer aconsejable llevar a cabo una reespecificación de la ecuación de comportamiento.

#### a) Resultados obtenidos en la detección del fraude en unidades monetarias

Los resultados obtenidos se sintetizan en el cuadro n.º 1 y en el gráfico 1.

Así, el cuadro n.º 1 separa entre los resultados obtenidos den-

CUADRO N.º 2

#### DETECCION DE DECLARACIONES FRAUDULENTAS

PROBABILIDAD SELECCIONADA	RESULTADOS DENTRO DE LA MUESTRA		RESULTADOS FUERA DE LA MUESTRA	
	% de la muestra inspeccionada (1)	% de declaraciones fraudulentas detectado (2)	% de la muestra inspeccionada (1')	% de declaraciones fraudulentas detectado (2')
P > 0,9	0,87	2,87	0,61	2,14
P > 0,8	1,26	3,72	0,92	2,99
P > 0,7	2,36	5,73	1,43	3,85
P > 0,6	3,78	9,46	2,56	7,26
P > 0,5	5,51	13,75	3,58	9,83
P > 0,4	8,89	20,92	5,62	14,53
P > 0,3	14,95	32,38	9,61	22,65
P > 0,2	32,49	58,74	21,57	44,44
P > 0,1	93,15	97,71	91,10	96,58
P > 0,0	100,00	100,00	100,00	100,00

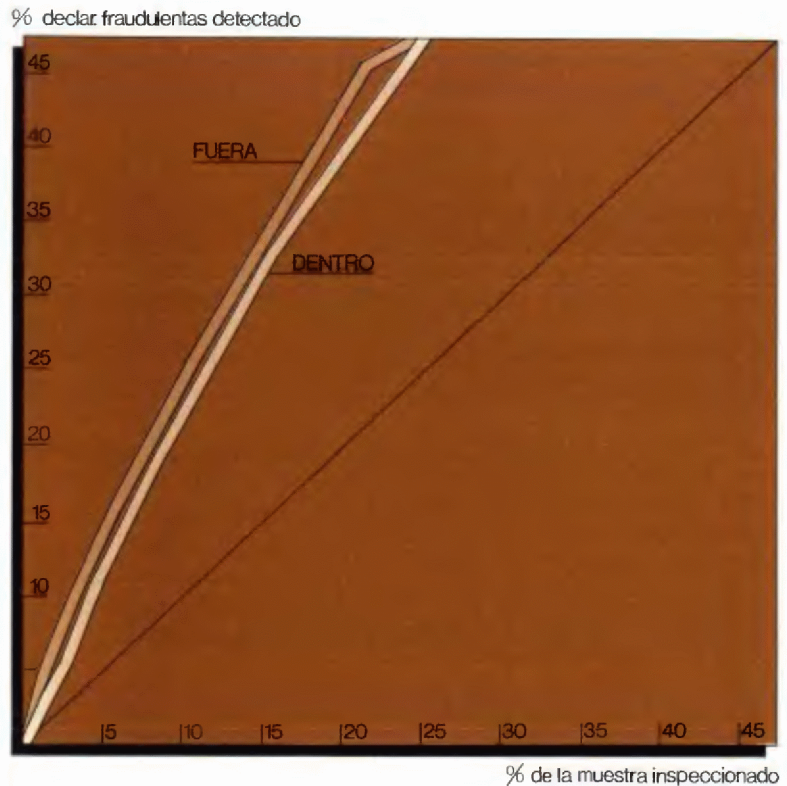
tro de la muestra de 1.270 observaciones y los que se obtienen aplicando el mismo modelo a una muestra de 978 observaciones ajena a la selección y estimación del modelo.

En efecto, dentro de la muestra, seleccionando un nivel de probabilidad de fraude en exceso a 0,9, el porcentaje de la muestra inspeccionada es de un 0,87 por 100. Este 0,87 por 100 de contribuyentes representa el 6,98 por 100 del fraude global. El *ratio* entre 6,98 y 0,87 da un valor de 8,03 y su diferencia con respecto a la unidad refleja la mejora que el modelo supone para este nivel de probabilidad frente a una selección de contribuyentes puramente aleatoria. Adicionalmente, el porcentaje de aciertos para esta probabilidad se sitúa en el 90,91 por 100. (Es decir, un 90,91 por 100 de los contribuyentes inspeccionados serían efectivamente fraudulentos.) Por último, dado que el porcentaje de declaraciones fraudulentas en la totalidad de la muestra se sitúa en el 27,45 por 100, el cociente entre ambos porcentajes, que da un valor de 3,31, refleja la mejora que el modelo supone en cuanto a la selección de contribuyentes fraudulentos con respecto a lo que sería una selección puramente aleatoria.

La interpretación de los resultados obtenidos fuera de la muestra es exactamente la misma. En este caso cabe resaltar que el porcentaje de declaraciones fraudulentas se sitúa en el 23,93 por 100.

Seleccionando un nivel de probabilidad —o punto de corte— razonable, como pueda ser una probabilidad de fraude en exceso a 0,6, la utilidad del modelo estriba en que inspeccionando

GRAFICO 2  
DETECCION DE DECLARACIONES  
FRAUDULENTAS



sólo, aproximadamente, un 2,5 por 100 de declaraciones, el porcentaje de fraude detectado se sitúa en el entorno del 12 por 100, obteniéndose un porcentaje de aciertos de alrededor del 70 por 100. Por otro lado, detectar alrededor de un 50 por 100 del fraude, comporta inspeccionar un 20 por 100 de la muestra. Posiblemente estos resultados podrán mejorarse. No obstante, la verdadera utilidad del modelo se da cuando la capacidad inspectora es limitada, en cuyo caso, seleccionando niveles de probabilidad elevados, el porcentaje de aciertos también lo es y la proporción del fraude

detectado se sitúa a un nivel muy superior al de la muestra inspeccionada.

Puede observarse que la capacidad de discriminación del modelo se mantiene dentro de unos límites razonables, al utilizar para su validación una muestra distinta de la empleada para su especificación y selección. El hecho de que los resultados empeoren ligeramente es plenamente explicable por el inevitable proceso de «agotamiento de los datos» que casi todo proceso de estimación econométrica entraña.

Por último, el gráfico 1 recoge estos resultados en forma de

curva de Lorenz. La capacidad de discriminación del modelo dentro de la muestra está reflejada a través de la línea de trazo continuo, y su capacidad de discriminación fuera de la muestra se indica por medio del trazo discontinuo. Puede observarse que la selección que el modelo sugiere se aparta claramente de la aleatoriedad, alternativa que vendría representada por la bisectriz, orientando la inspección hacia las declaraciones fraudulentas.

**b) Resultados obtenidos en la detección de declaraciones fraudulentas**

En el cuadro y en el gráfico precedentes se analiza el sesgo que el modelo introduce en la detección del fraude en pesetas, frente a lo que sería una selección puramente aleatoria. En este apartado se repite el mismo análisis, pero centrando la atención en la capacidad que el modelo tiene para detectar declaracio-

nes fraudulentas, sin tomar en consideración el importe del fraude en tales declaraciones.

Así, el cuadro n.º 2 señala que, para una probabilidad de fraude en exceso a 0,9, dentro de la muestra el modelo selecciona un 0,87 por 100 de declaraciones, que representan el 2,87 por 100 del total de declaraciones fraudulentas. Fuera de la muestra, a este mismo nivel de probabilidad, un 0,61 por 100 de declaraciones seleccionadas permite detectar el 2,14 por 100 de declaraciones fraudulentas. Para una probabilidad de fraude en exceso de 0,6, aproximadamente, un 2 por 100 de la muestra inspeccionada detecta del orden de un 7 por 100 del total de declaraciones fraudulentas, y, como se vio en el apartado precedente, tales declaraciones representan un 12 por 100 del fraude global en pesetas.

El gráfico 2 establece la correspondiente representación gráfica. A este respecto se com-

prueba que el comportamiento del modelo dentro de la muestra es similar al que se obtiene al utilizar una muestra ajena a la empleada para su selección y estimación.

**c) Resultados obtenidos en la predicción del fraude agregado**

A partir del conjunto de declaraciones de renta presentadas en un determinado ejercicio, pueden suscitarse dos cuestiones básicas. En primer lugar, ¿cuál es el fraude agregado o global que cabe estimar para el conjunto de contribuyentes que han presentado la correspondiente declaración sobre la renta? En segundo lugar, ¿cuál es la predicción del fraude global que se detectará al inspeccionar distintos porcentajes de la muestra según la ordenación de declaraciones que el modelo establece? En este sentido, el cuadro n.º 3 ofrece la información relevante

CUADRO N.º 3

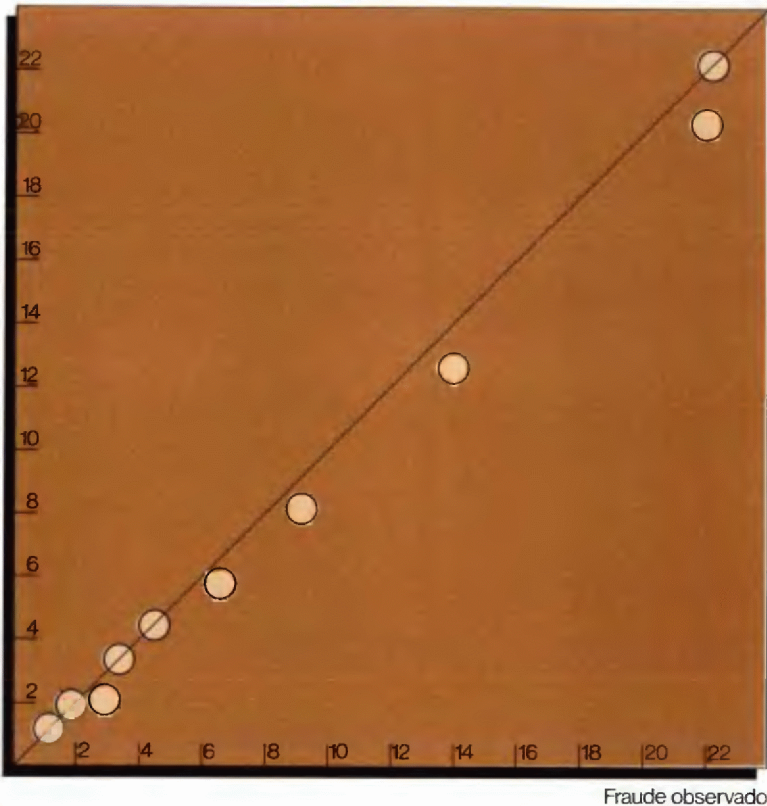
**PREDICCIÓN DEL FRAUDE AGREGADO**

PROBABILIDAD SELECCIONADA	RESULTADOS DENTRO DE LA MUESTRA				RESULTADOS FUERA DE LA MUESTRA			
	% de la muestra inspeccionada	Fraude predicho (miles)	Fraude observado (miles)	Porcentaje de error	% de la muestra inspeccionada	Fraude predicho (miles)	Fraude observado (miles)	Porcentaje de error
P > 0,9	0,87	1.417	1.565	- 9,46	0,61	1.081	821	+ 31,67
P > 0,8	1,26	1.694	1.929	- 12,18	0,92	1.265	1.139	+ 11,06
P > 0,7	2,36	2.479	2.416	+ 2,61	1,43	1.559	1.203	+ 29,59
P > 0,6	3,78	3.341	3.350	- 0,27	2,56	2.130	1.792	+ 18,86
P > 0,5	5,51	4.314	4.334	- 0,46	3,58	2.590	2.310	+ 12,12
P > 0,4	8,89	5.986	6.231	- 3,93	5,62	3.359	3.106	+ 8,14
P > 0,3	14,95	8.237	9.021	- 8,69	9,61	4.623	4.364	+ 5,93
P > 0,2	32,49	12.559	14.141	- 11,19	21,57	7.137	7.627	- 6,42
P > 0,1	93,15	20.229	21.952	- 7,85	91,10	14.159	14.846	- 4,63
P > 0,0	100,00	22.407	22.407	0,00	100,00	14.447	15.309	- 5,63

Media de los valores absolutos de los porcentajes de error. 5,66 % 13,40 %  
 Raíz cuadrada del error cuadrático medio 7,23 % 16,43 %

**GRAFICO 3  
PREDICCIÓN DEL FRAUDE AGREGADO:  
RESULTADOS OBTENIDOS DENTRO  
DE LA MUESTRA**

Fraude predicho



a efectos de la evaluación de la capacidad predictiva del modelo.

Con respecto a la primera cuestión, y analizando el comportamiento del modelo fuera de la muestra que ha servido para su selección y estimación (en este caso, tratar de evaluar el comportamiento del modelo dentro de la muestra carece de sentido dado que los residuos suman cero), a partir de la aproximación «Tobit» el fraude global predicho es de 14.447 miles de pesetas. El fraude global observado es de 15.309 miles de pesetas. En consecuencia, el modelo hubiese in-

fravalorado en un 5 por 100 el volumen global de fraude. Si tal análisis se hace extensible a la totalidad de declaraciones de renta de un determinado ejercicio, el modelo puede ofrecer una aproximación al fraude global que cabe estimar para el conjunto de contribuyentes que han presentado declaración. En este caso, no obstante, la estabilidad del modelo es más discutible, al analizar ejercicios distintos del que ha servido para su estimación, dado que el porcentaje de fraude puede variar en el tiempo.

Con respecto a la segunda cuestión, el modelo señala que, para un nivel de probabilidad de fraude en exceso, por ejemplo, a 0,6, dentro de la muestra la inspección de un 3,78 por 100 de declaraciones permitiría detectar 3.341 miles de pesetas de fraude. El valor observado es en realidad de 3.350 miles de pesetas. Fuera de la muestra, el modelo predice que inspeccionando un 2,56 por 100 de declaraciones, aflorarían 2.130 miles de pesetas de fraude. El valor observado es de 1.792 miles de pesetas. En cuanto a la representación gráfica de estos resultados se efectúa en los gráficos 3 y 4.

En cualquier caso, conviene resaltar que esta predicción del fraude global puede realizarse para cualquier ordenación de declaraciones. A título ilustrativo, para una ordenación de declaraciones llevada a cabo por la propia inspección atendiendo a otros criterios distintos de la probabilidad de fraude que el modelo predice.

Desde la óptica de la predicción del fraude global, los resultados del modelo son aceptables, sobre todo cuando tal evaluación hace referencia a un colectivo de declaraciones no excesivamente reducido.

#### **IV. CONCLUSIONES**

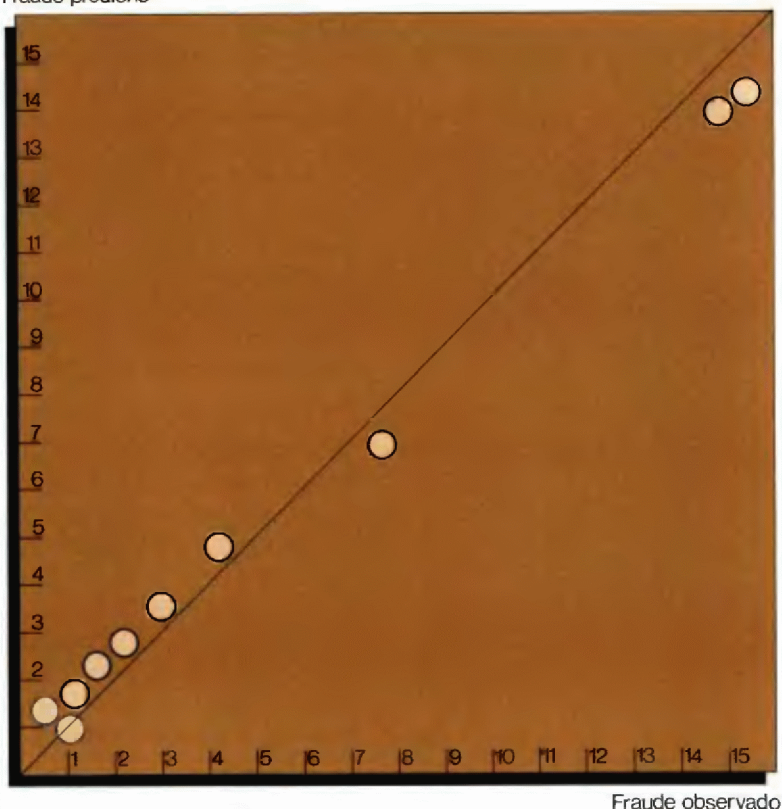
Como conclusiones de este trabajo pueden destacarse los siguientes aspectos:

a) En primer lugar, cabe resaltar la utilidad de disponer de un modelo estadístico de ordenación de declaraciones según la probabilidad de fraude al tiempo que de predicción de su importe, dado que en tal caso la



**GRAFICO 4  
PREDICCIÓN DEL FRAUDE AGREGADO:  
RESULTADOS OBTENIDOS FUERA DE LA  
MUESTRA**

Fraude predicho



Fraude observado

bien cabe esperar que estos resultados puedan mejorarse en el curso de la investigación. Por otro lado, la capacidad predictiva del fraude agregado que se deduce del modelo se sitúa en unos márgenes de error aceptables.

c) Dentro de las mejoras potenciales del modelo, debe hacerse hincapié en los siguientes puntos:

En primer lugar, llevar a cabo su estimación por métodos más adecuados. En la actualidad se está ultimando la aplicación de Máxima Verosimilitud para la estimación de los modelos «Probit» y «Tobit». No obstante, los resultados que en este artículo se ofrecen hacen sólo referencia a la aplicación de los modelos Logit y de Regresión por Mínimos Cuadrados, y, en consecuencia, son inferiores a los que se obtendrán al emplear otros métodos de estimación más eficientes.

En segundo lugar, puede tener interés contemplar la posible inclusión de otras variables explicativas que en esta estimación no han podido ser tratadas por falta de información disponible. Se da tal circunstancia, a título ilustrativo, para el caso de información referente a la declaración sobre el patrimonio, al tratamiento más detallado de las profesiones, a la ubicación geográfica de los contribuyentes, o a datos relativos a la acción inspectora a que el contribuyente ha sido sometido con anterioridad.

Por último, cabría tratar de aplicar el modelo a nivel de distintas profesiones, permitiendo cambios estructurales en la totalidad o parte de los coeficientes al pasar de una a otra profesión.

totalidad del proceso es plenamente informatizable. En este sentido, a todas las declaraciones se les aplica igual baremo a efectos de llevar a cabo su ordenación, si bien ello no excluye la posibilidad de recurrir a otros criterios complementarios con el fin de establecer la selección definitiva de declaraciones a inspeccionar.

b) Los resultados provisionales del modelo son esperanzadores, en el sentido de que, al efectuar su evaluación con una muestra distinta de la utilizada para la especificación y estima-

ción, se comprueba que inspeccionando declaraciones con elevada probabilidad de fraude, lo que comporta seleccionar un reducido porcentaje del total de la muestra, el porcentaje de aciertos (o clasificaciones correctas) así como el porcentaje de fraude detectado, son magnitudes que se sitúan muy por encima de lo que comportaría una selección aleatoria. Al reducirse el nivel de probabilidad que orienta la selección de declaraciones, el porcentaje de la muestra inspeccionada aumenta y, como es lógico, la capacidad de discriminación del modelo empeora, si

En conjunto, todo ello debe contribuir a mejorar la eficacia de la actividad inspectora cuando los medios de que dispone la Administración son limitados. Se trata de asignar estos medios de forma rentable desde la óptica de la detección del fraude, de suerte que se maximice el *ratio* entre volumen de fraude detectado y costes de la inspección.

#### NOTA

(\*) Este trabajo constituye un resumen del estado actual de una investigación financiada por el Instituto de Estudios Fiscales. Recientemente se ha incorporado al equipo investigador el profesor Jaime García, de la Universidad Autónoma de Barcelona.

Deseamos expresar nuestro agradecimiento al Instituto de Estudios Fiscales, así como al Centro de Proceso de Datos del Ministerio de Economía y Hacienda, por la colaboración prestada. Asimismo, agradeceremos a Antonio Zabalza sus comentarios y sugerencias a este trabajo. No obstante, asumimos plenamente la responsabilidad de los posibles errores.

#### REFERENCIAS

- AMEMIYA (1981), «Qualitative Response Models: A Survey», *Journal of Economic Literature*, 19, 4, págs. 483-536.
- MC. FADDEN (1973), «Conditional Logit Analysis of Qualitative Choice Behavior», en ZAREMBKA *et al.*, *Frontiers in Econometrics*, Academic Press.
- (1976), «Quantal Choice Analysis: A Survey», *Annals of Economic and Social Measurement*, 5, págs. 363-390.
- MADDALA (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.
- VALDES *et al.* (1982), «Los Métodos de Análisis Discriminante como Herramienta al Servicio de la Inspección Fiscal», Instituto de Estudios Fiscales, Monografía, n.º 21.