QÜESTIIÓ, vol. 22, 2, p. 253-289, 1998

REFLEXIONES SOBRE LA ESTRATEGIA DE MEDIDA DE LOS CAMBIOS EN PROBABILIDAD EN MODELOS DE ELECCIÓN BINARIOS

M.T. APARICIO*
I. VILLANÚA*
Universidad de Zaragoza⋆

Este trabajo se centra en la evaluación de la medida que, en el marco de los modelos de elección binarios o dicotómicos, se utiliza para reflejar el cambio en la probabilidad ante la variación de una de las variables explicativas. La opción de cuantificación más común ha consistido en utilizar el vector de valores medios de las variables explicativas, lo que podemos entender como poner el énfasis en el comportamiento de «*un individuo medio*»*. Frente a esta práctica habitual, efectuamos una propuesta de evaluación alternativa que atiende fundamentalmente al* «*comportamiento medio de la muestra*»*. Se diseña, asimismo, un experimento de Monte Carlo que permite probar que la propuesta desarrollada en este trabajo es más robusta que la tradicionalmente planteada en la literatura.*

Reflections on the probability change measurement strategy in binary choice models.

Palabras clave: Elección discreta, variación de probabilidad, robustez.

Clasificación AMS: 62J02, 65C05

* Los autores desean agradecer los comentarios de Antonio Aznar, Carmen García-Olaverri y de un evaluador anónimo, así como la financiación de DGICYT PB94-0602.

⋆ Departamento de Análisis Económico. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza. Gran Vía, 2. 50005 Zaragoza.

– Recibido en julio de 1996.
– Aceptado en enero de 1998.

1. INTRODUCCIÓN

En los últimos años, los denominados modelos de elección discreta (MED) han tenido un auge considerable, debido al mayor énfasis de la investigación econométrica en modelos desagregados. Este hecho ha venido propiciado por dos razones fundamentales; en primer lugar, el reconocimiento de que el comportamiento económico relevante debe ser tratado a nivel individual, mejor que mediante la consideración de agregados de acciones individuales y, en segundo lugar, por la posibilidad cada vez mayor de disponer de datos a nivel microeconómico, es decir, de las unidades que toman las decisiones.

El término MED hace referencia a toda una clase de modelos que tratan de calcular la probabilidad de que una unidad de decisión escoja una alternativa concreta de entre un conjunto de opciones, dada la respuesta del individuo sobre la elección efectuada (y) y todo un conjunto de variables socio-económicas (x) de las cuales depende obviamente la elección y que representan características de las alternativas disponibles y características del sujeto decisor. Se pueden distinguir diferentes modelos dentro de la clase mencionada, atendiendo a la diferente forma funcional que relaciona los datos observados con la probabilidad o bien según la extensión del conjunto de elección; en este último sentido, el planteamiento más sencillo corresponde a los MED dicotómicos. Centrándonos en estos, la probabilidad de que la unidad de decisión $i-$ésima elija una de las alternativas se especifica como una función general de la forma:

$$p_i = F(x_i'\beta) \tag{1}$$

donde x_i es el vector de variables socio-económicas continuas y discretas asociadas con la unidad $i-$ésima y β es un vector de parámetros desconocidos. La justificación teórica de esta formulación se establece en base a la naturaleza de los procesos de toma de decisiones, utilizando los conceptos de la teoría de la utilidad como puede verse, entre otros, en Amemiya (1981), McFadden (1970), Maddala (1983) o Train (1986).

A partir de (1), la determinación concreta de F conduce a modelos específicos como el modelo probit o el modelo logit, según se identifique F con la función de distribución de la normal tipificada o de la logística estandarizada, respectivamente. En ambos casos, la estimación de los coeficientes se realiza por el método de máxima verosimilitud, requiriéndose el uso de algoritmos de optimización para modelos no lineales; tal estimación permite obtener, para cada unidad de decisión, la probabilidad estimada de elección.

A pesar del gran número de trabajos empíricos que utilizan este tipo de modelos, la etapa de validación de los mismos suele recibir bastante poca atención y lo normal es presentar los coeficientes estimados y sus correspondientes $t-$ratios, acompañados de una de las múltiples medidas escalares de bondad del ajuste propuestas para planteamientos de este tipo[1] y del contraste LR relativo a la significación conjunta del modelo. La interpretación de las magnitudes anteriores conforman las conclusiones del estudio, las cuales se completan frecuentemente con el análisis del efecto que sobre la probabilidad de elección tienen las variaciones de las diferentes variables socio-económicas que intervienen en el supuesto concreto. Esta forma tan escueta de entender la etapa de evaluación del modelo se refleja claramente en la afirmación de Pagan y Vella (1989), cuando en relación con los estudios que utilizan MED establecen: «Un pequeño muestreo de la literatura revela una gran diferencia respecto de la vieja tradición de los modelos de series temporales, en el sentido de no prestar gran atención a la calidad del modelo estimado».

El objetivo de este trabajo es centrarse en una de las cuestiones que hemos mencionado como integrantes de la valoración habitual de este tipo de modelos. En concreto, nos interesa profundizar en la medida utilizada para determinar la variación de la probabilidad ante cambios de las variables explicativas. Como es bien conocido, en este tipo de modelos no lineales la magnitud de los coeficientes por sí sola no permite sintetizar tal información y el instrumento adoptado a tal fin exige calcular $x_i\hat{\beta}$, con lo cual, una vez obtenido el vector de estimadores habrá que concretar el valor que se adopta para las diferentes variables explicativas.

El análisis de las diferentes estrategias de evaluación, habitualmente propuestas en la literatura, constituye el núcleo central de este trabajo, de modo que la discusión acerca de las ventajas e inconvenientes de cada una de ellas nos sirve de base para proponer una vía alternativa de evaluación que hace hincapié en el comportamiento medio de la muestra y que, desde nuestro punto de vista, refleja un mejor uso de la evidencia muestral.

La comparación de la estrategia propuesta con la práctica habitual se realiza diseñando un experimento de simulación que permita analizar la «robustez» de la nueva medida.

Establecido el tema objeto de discusión, el trabajo se estructura de la forma siguiente: en la Sección 2 nos centramos en el instrumento que refleja la variación en la probabilidad, con la única intención de tratar de rebatir algunas interpretaciones del mismo que consideramos erróneas y que han llevado a algunos autores a cuestionar su validez; en la Sección 3 establecemos las diversas estrategias existentes para concretar la variación mencionada poniendo de manifiesto las limitaciones que, a nuestro juicio, contienen lo que nos lleva a proponer un esquema amplio de razonamiento alternativo; en la Sección 4 se lleva a cabo un ejercicio de simulación para determinar posibles pautas de comportamiento de la estrategia propuesta en relación con el

[1]Medidas de este tipo, denominadas comúnmente como pseudo-R^2 han sido propuestas por McFadden (1974), Maddala (1983), Aldrich y Nelson (1984) y McKelvey y Zavoina (1975). Una comparación de todos ellos es presentada en Windmeijer (1995).

enfoque habitual en la literatura; en la Sección 5 se ilustran las distintas estrategias de análisis y su comportamiento mediante un ejemplo con datos reales, y en la última se resumen las conclusiones principales.

2. LA VARIACIÓN DE LA PROBABILIDAD ANTE CAMBIOS DE LAS VARIABLES INDEPENDIENTES

Como es bien conocido, en un modelo econométrico lineal la magnitud de los coeficientes que acompañan a las variables explicativas, proporcionan la variación de la variable endógena ante un cambio unitario de las primeras; asimismo, el signo del coeficiente está indicando la dirección de tal cambio. En consecuencia, los coeficientes son interpretables directamente, tanto en magnitud como en signo.

En el caso de los modelos no lineales que nos ocupan, lo que se predice es la probabilidad de elección de una alternativa, de modo que el efecto de interés se centra en la variación de tal probabilidad con respecto a la variable explicativa; pero dada la forma funcional específica del modelo, la magnitud del coeficiente no nos da directamente este efecto, aunque el signo del coeficiente seguirá reflejando el sentido del cambio.

Lo que interesa, en definitiva, es obtener el ratio entre la variación de la probabilidad y la propia variación de dicha variable, es decir, $\Delta p_i/\Delta x_j$ cuya expresión analítica podemos escribir como:

$$\frac{\Delta p_i}{\Delta x_j} = \frac{F\left(x_i'\beta + \beta_j \Delta x_j\right) - F\left(x_i'\beta\right)}{\Delta x_j} \tag{2}$$

de modo que para variaciones infinitesimales de x_j resulta:

$$\lim_{\Delta x_j \to 0} \frac{\Delta p_i}{\Delta x_j} = \lim_{\Delta x_j \to 0} \frac{F\left(x_i'\beta + \beta_j \Delta x_j\right) - F\left(x_i'\beta\right)}{\Delta x_j} = \frac{\partial p_i}{\partial x_j} \tag{3}$$

concretando $\partial p_i/\partial x_j$ para los modelos expresados en (1) se tiene:

$$\frac{\partial p_i}{\partial x_j} = f\left(x_i'\beta\right)\beta_j \tag{4}$$

donde $f(\cdot)$ denota la correspondiente función de densidad. La expresión (4) muestra con claridad lo manifestado anteriormente, en el sentido de que β_j por sí solo no indica la magnitud de la variación, la cual, además, depende de los valores concretos adoptados por las variables explicativas, mientras que el signo del coeficiente sí que determina la dirección del cambio, puesto que $f(\cdot)$ siempre será positiva por ser una función de densidad.

La consideración de la igualdad reflejada en (3) permite una interpretación correcta de la derivada parcial en modelos de este tipo (no lineales), y evita confusiones como las reflejadas en los artículos de Petersen (1985) y LeCleer (1992). Ambos autores critican la utilización de (4) por no ser una medida acotada entre cero y uno y, en consecuencia, posibilitar la generación de probabilidades fuera del rango plausible. En este sentido, LeCleer (1992) argumenta: «Una interpretación común pero inapropiada de la derivada parcial evaluada en la media de las variables independientes, es como probabilidad marginal con respecto a un cambio unitario en cada variable independiente. El problema está en que la derivada parcial no tiene límite y puede generar probabilidades fuera del rango plausible».

Por su parte Petersen (1985) apunta: «La fórmula frecuente e incorrectamente utilizada para calcular el cambio en p_i ante un cambio en x_j es $\partial p_i/\partial x_j$. Pero esto no nos da el cambio en p_i por un cambio unitario en x_j, ni tampoco por un cambio infinitesimal en x_j». El autor justifica estas afirmaciones diciendo que la derivada en funciones no lineales tiene una interpretación diferente, en el sentido de proporcionarnos la pendiente de p_i con respecto a x_j, cuando el denominador se aproxima a cero, y concluye afirmando: «… la pendiente no tiene límite porque los coeficientes no están acotados. Así pues, la pendiente puede ser mayor que uno y generar probabilidades fuera del intervalo cero-uno».

Desde nuestro punto de vista, no existe ningún problema en la utilización de (4) si, como hemos dicho antes, entendemos lo que realmente significa, esto es, el cambio en p_i en proporción al de x_j, de modo que para variaciones pequeñas de x_j, el cambio en probabilidad se aproximará a $\frac{\partial p_i}{\partial x_j}\Delta x_j$. En cualquier caso, la variación de x_j se considerará grande o pequeña dependiendo del rango de valores en el que se mueve la variable explicativa.

LeCleer (1992) ilustra lo que considera el problema de la no acotación, utilizando datos reales obtenidos de un estudio realizado por Hill y Ingram (1989) sobre la elección por parte de las entidades de préstamos entre dos principios contables, mediante la aplicación de un modelo logit. En base a la información suministrada, via cuestionario, por 113 entidades financieras de Estados Unidos, correspondiente al primer semestre de 1981, sobre un conjunto de variables financieras (activo total, activo neto/depósitos, préstamos hipotecarios/activo, depósitos/activo, etc.) se formulan el conjunto de variables explicativas que aparecen en el modelo logit objeto de estimación. Utilizando los resultados de esta etapa y el conjunto de medidas descriptivas que, de cada factor, proporcionan los autores, LeClere (1992) calcula $\partial p_i/\partial x_j$ para cada variable independiente de tipo continuo, obteniendo para una de ellas un valor de 89,0082 (para el resto de variables continuas, $\partial p_i/\partial x_j$ se encuentra entre cero y uno), lo que le lleva a cuestionar la consideración de la expresión (4) como media del cambio en probabilidad.

Ahora bien, si efectuamos una interpretación correcta de la derivada considerando como unidad de cambio no la unidad en sí, sino lo que podemos denominar como «unidad de cambio coherente», esto es, la correspondiente al intervalo en que se mueve la variable, la utilización de este instrumento no plantearía ningún problema. En este caso concreto, las medidas descriptivas correspondientes a la variable en cuestión proporcionan una media de $0'0049$, una desviación típica de $0'0028$ y un rango de valores entre 0 y $0'0209$, todo lo cual indica claramente que la «unidad de cambio coherente» no va a ser la unidad y, en consecuencia, la variación en la probabilidad ante cambios unitarios (en el sentido mencionado) de dicha variable, estará ya comprendida entre cero y uno como corresponde a tal concepto.

Podemos concluir, por tanto, que cuando estamos analizando el cambio en probabilidad ante un cambio unitario (entendiendo la unidad en el contexto correspondiente) de la variable independiente, la expresión analítica que se aproxima a dicho cambio es $\frac{\partial p_i}{\partial x_j}\Delta x_j$ y esta aproximación será mejor a medida que $\Delta x_j \to 0$; de esta manera, la no acotación de la derivada no plantea ningún problema.

Todo el análisis efectuado exige que las variables x's sean continuas, por cuanto la derivada parcial no existe cuando tales variables son de tipo discreto. Ahora bien, en la mayoría de las especificaciones que nos ocupan, alguna o varias de las variables que representan las características socio-económicas son variables discretas, normalmente variables ficticias con las que se intentan representar factores no cuantificables. Para tales variables, la evaluación del cambio en probabilidad debe hacerse teniendo en cuenta los cambios discretos de dicha variable y lo que interesa es obtener:

$$\frac{\Delta p_i}{\Delta x_j} = \frac{P(y_i = 1|x_j = 1) - P(y_i = 1|x_j = 0)}{1 - 0} \tag{5}$$

en otros términos, realizar un análisis de sensibilidad, calculando p_i para cada uno de los posibles valores adoptados por la variable discreta y obteniendo mediante su diferencia el cambio en la probabilidad de elección.

3. ESTRATEGIAS DE ANÁLISIS

En el caso de variables continuas, hemos aceptado que la interpretación de la derivada como razón entre la variación de p_i y la variación de x_j, cuando Δx_j tiende a cero, permite establecer que la variación absoluta de p_i vendrá expresada, aproximadamente, mediante $\frac{\partial p_i}{\partial x_j}\Delta x_j$ y tomando como «unidad de cambio» de x_j una unidad finita, pero coherente con el rango plausible, en los términos definidos anteriormente, de valores de la correspondiente variable explicativa, la falta de acotación de la derivada no supone ningún problema. Cuando las variables explicativas son de tipo discreto, la variación de p_i se establece llevando a cabo lo que hemos denominado en el apartado anterior, un análisis de sensibilidad.

En cualquiera de las dos situaciones, la concreción de la variación en la probabilidad exige calcular $x_i'\beta$. El vector de parámetros se habrá estimado previamente, pero necesitamos decidir el valor que adoptamos para el conjunto de las variables explicativas y a este respecto no existe una estrategia única, pudiéndose encontrar en la literatura diferentes formas de actuación que pasamos a describir y analizar.

La opción más común consiste en efectuar la evaluación en el vector de valores medios de las diferentes variables explicativas, como puede verse, entre otros, en LeCleer (1992), Greene (1993) y Greene Knapp y Seaks (1992). En nuestra opinión, esta forma de actuación no resulta del todo adecuada debido, fundamentalmente, a dos hechos. En primer lugar, la evaluación del conjunto de variables explicativas en su valor medio, cuando como tales suele ser corriente que aparezcan variables ficticias, resulta incoherente con el sentido que se les otorga a dichas variables. La media en tal caso simplemente establece la proporción de una de las características sobre el total, pero no es un valor que la variable ficticia pueda adoptar y que tenga significado en dicho contexto. Por otra parte, utilizar la media muestral para las variables de tipo contínuo, puede plantear problemas de falta de representatividad en el caso de que los valores de una determinada variable estén muy dispersos. En cualquier caso, la cuestión relevante es el tratamiento eficiente de la información y, a este respecto, parece más adecuado utilizar el conjunto de datos relativos a cada unidad de decisión —obteniendo $\partial p_i/\partial x_j$ para cada una de las unidades de la muestra— y resumir el resultado en un solo valor, calculando entonces el valor medio. Se trata, en definitiva, de poner el énfasis en el comportamiento medio de la muestra antes que en el comportamiento de un «individuo medio», al objeto de evitar una pérdida de información sin ningún tipo de justificación. Esta postura parece ser también defendida por Train (1986), cuando plantea la inconsistencia de calcular la derivada en la media en lugar de la media de las derivadas, basándose en la no coincidencia, para modelos no lineales, entre la media de una función y el valor de la función en el punto medio.

No obstante, tal alternativa de evaluación no es posteriormente utilizada por el autor y su propuesta, al igual que la de otros autores como Jhonson (1987) y Maddala (1992), consiste en calcular $\partial p_i/\partial x_j$ para aquellos valores de las variables explicativas que sean más significativos, con el fin de obtener una idea sobre el rango de variación de los cambios resultantes en la probabilidad. Esta opción adolece de cierta ambigüedad por cuanto no se explicita lo que se consideran «valores más significativos» de las explicativas, ¿serían los valores de las x's más frecuentes en la muestra concreta?, ¿nos referimos a los valores máximo y mínimo de cada variable y sus posibles combinaciones? Train (1986) discute este aspecto teniendo en cuenta que el valor de la derivada será máximo para $p_i = 0'5$ (o equivalentemente, de acuerdo con (1), $x_i'\beta = 0$, lo cual, si el modelo no contiene término independiente determina que todas las explicativas deben ser igual a cero y caso de existir término autónomo dará lugar a diferentes combinaciones de valores de x que satisfagan $x_i'\beta = 0$) y mínimo cuando $p_i = 0$ y $p_i = 1$ (lo que se corresponde respectivamente con $x_i'\beta = -\infty$ y $x_i'\beta = +\infty$, situaciones que sólo tienen interés como acotación, pero que son inalcanzables en la práctica) de tal modo que son estos tres puntos los que constituyen la propuesta de evaluación. El interés de esta estrategia como instrumento para informar sobre el comportamiento medio de la muestra es bastante escaso, puesto que se limita a establecer, exclusivamente, el rango de variación de la probabilidad tomando como referencia los puntos límite.

Fisher (1991) plantea una variación de esta última propuesta basada en la utilización de la media de la variable dicotómica de respuesta observada (y). En concreto, tal valor le permite, invirtiendo (1) de modo que $\widetilde{x_i'\beta} = F^{-1}(\bar{y})$, obtener la magnitud de $x_i'\beta$ a utilizar para evaluar la derivada. La idea de este autor resulta difícilmente justificable, por cuanto no considera el conjunto de información sintetizado en las variables explicativas del modelo, y en su lugar identifica a $\bar{y}$ como una estimación de la probabilidad de elección, lo que equivale a obviar por completo el marco teórico de razonamiento que subyace en la especificación de los MED.

Podemos decir, por tanto, que una característica común al conjunto de alternativas de evaluación analizadas, es la falta de un buen tratamiento de la información que permita dar cuenta del comportamiento medio de la muestra. En consecuencia, al objeto de plantear una vía adecuada para medir la variación en la probabilidad ante cambios de las variables explicativas, nuestra propuesta consiste en calcular la derivada para cada elemento de decisión y plasmar, posteriormente, el resultado en el valor medio de tales derivadas. De esta manera, se está utilizando el conjunto de información muestral del que realmente se dispone y se evitan las posibles inconsistencias asociadas con un uso parcial de la información.

Hay que tener presente, no obstante, que esta propuesta de análisis solo es válida para variables de tipo continuo, de manera que el estudio referido a las posibles variables ficticias que integren el conjunto de factores explicativos deberá hacerse utilizando la expresión (5). El cálculo de dicha expresión ha consistido, normalmente, en fijar las variables continuas en el nivel asociado con el valor medio de las mismas. En nuestra opinión, y de acuerdo con la propuesta de estudiar el «comportamiento medio de la muestra», nos parece más coherente obtener la probabilidad p_i para cada elemento de decisión y calcular la media para cada submuestra concreta distinguida por el valor asociado a la variable ficticia, de modo que la diferencia entre estas medias nos dará el cambio en probabilidad ante un cambio en el valor de la variable ficticia.

La metodología propuesta para analizar la variación de la probabilidad ante cambios en las variables explicativas se fundamenta, por tanto, en la idea de un uso eficiente de la información muestral disponible que permita dar cuenta, de manera coherente, del comportamiento medio de la muestra.

4. EJERCICIO DE SIMULACIÓN

En el apartado anterior hemos revisado, en primer lugar, las distintas estrategias relativas a la concreción de la variación de la probabilidad y los problemas que, en nuestra opinión llevan asociados, lo cual nos ha permitido justificar, posteriormente, una opción alternativa.

Al trabajar en el contexto de funciones no lineales, se verifica que la media de una función de este tipo no va a coincidir con la función evaluada en el punto medio; en consecuencia, la alternativa de actuación propuesta en este trabajo, fundamentada en el cálculo de la media de las derivadas, no va a ser igual a la comúnmente utilizada, basada en la derivada en el punto medio. Por ello, resulta de interés efectuar un análisis de ambas medidas al objeto de poder determinar, si es posible, la superioridad de una frente a la otra de acuerdo con algún tipo de criterio.

Un aspecto previo a la cuestión central mencionada consistirá en estudiar si la diferencia que va a producirse entre ambas estrategias tiene, en la mayoría de los casos, el mismo sentido o, por el contrario, no responde a ninguna pauta fija. En otros términos, se intenta verificar si existe un posible patrón de comportamiento para las dos alternativas planteadas, lo cual, aunque no permite establecer preferencias entre ellas, sí que parece, en el caso de encontrarse, un resultado más satisfactorio.

Con este objeto se ha diseñado un ejercicio de simulación que nos permitirá, utilizando como proceso generador de los datos (PGD) bien un modelo probit o bien un modelo logit, generar diferentes muestras. En concreto, generaremos 100 muestras —es decir, se efectúan 100 replicaciones— de tamaño 100 cada una de ellas.

El procedimiento de generación está basado en los trabajos de Gourieroux, Monfort, Renault y Trognon (1987) y Griffiths, Carter y Pope (1987). En el primero de ellos se adopta como punto de partida la justificación teórica de los MED y, más concretamente, la relación lineal que se establece entre el índice subjetivo (no observable) denominado de utilidad o de satisfacción correspondiente al individuo $i-$ésimo (y_i^*) y el conjunto de características socio-económicas que influyen en el problema concreto, pudiendo escribir:

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

de modo que si la función de distribución asumida para u_i es $N(0,1)$ especificaremos, en última instancia, un modelo probit y caso de suponer una distribución logística resultará un modelo logit.

Suponiendo que el vector x_i', además del término independiente, está integrado por dos variables que suponemos cuantitativas, y que denotaremos por x_1 y x_2, el proceso seguido consiste en generar éstas como $N(0,1)$, de igual modo que el término de perturbación y fijar los parámetros igual a la unidad resultando, por tanto, un primer «índice subjetivo» dado por:

$$y_i^* = 1 + x_{1i} + x_{2i} + u_i$$

De esta manera, el conocimiento del conjunto de observaciones de y_i^*, x_{1i} y x_{2i}, permite estimar por máxima verosimilitud el modelo:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \tag{6}$$

A partir de esta expresión, utilizando el vector de coeficientes estimados y una muestra de tamaño 100 de x_1, x_2 y del término de perturbación (que también hemos asumido $N(0,1)$), generamos la correspondiente muestra de tamaño 100 para el índice subjetivo (y_i^*). Como lo que realmente interesa es el conocimiento de la respuesta del individuo, esto es, la obtención de la variable dicotómica representada como y_i, deberemos hacer uso de la conocida relación entre y_i^* e y_i que podemos concretar como:

$$y_i = \begin{cases} 1 & \text{si} \quad y_i^* > 0 \\ 0 & \text{si} \quad y_i^* \leqslant 0 \end{cases}$$

y, de este modo, obtenemos una muestra de tamaño 100 para y_i. Este proceso, tal como se ha mencionado anteriormente, se repite 100 veces, es decir, establecemos el número de réplicas igual a 100.

En el segundo de los trabajos citados se plantea una forma alternativa de generación de los datos según la cual, fijados los parámetros y las variables explicativas, de acuerdo con los mismos supuestos establecidos anteriormente, se obtiene $p_i = F(x_i'\beta)$. A continuación se genera una variable uniforme en el intervalo (0,1) que denominamos como v_i y la variable dicotómica se obtiene de acuerdo con:

$$y_i = \begin{cases} 1 & \text{si} \quad v_i \in [0, P_i] \\ 0 & \text{si} \quad v_i \in (P_i, 1] \end{cases}$$

De este modo, la consideración de diferentes variables v_i permite obtener las diversas muestras de la variable dicotómica.

Griffiths, Carter y Pope (1987) desarrollan este procedimiento en el marco del modelo probit, aunque su aplicación para el modelo logit resulta inmediata, por cuanto sólo requiere identificar F con la función de distribución logística en la obtención de p_i.

En la generación de los modelos probit hemos aplicado el primero de los procedimientos expuestos comprobando, adicionalmente, que las conclusiones obtenidas no variaban cuando tales modelos eran generados de acuerdo con el segundo de los métodos. Esta similitud nos lleva a utilizar el planteamiento de Griffiths, Carter y Pope (1987), que requiere menor complejidad de cálculo, para la generación de los modelos logit. En ambos casos, el programa informático utilizado ha sido Gauss.

Descrito el proceso de generación de los diferentes modelos, podemos pasar a verificar la cuestión planteada en cuanto a la existencia o no de un patrón de comportamiento estable entre las dos medidas objeto de análisis. Denominando $D_K \quad (K = 1,2)$ a la derivada en el punto medio y $DM_K \quad (K = 1,2)$ a la media de las derivadas, las tablas 1 y 2 del Apéndice presentan ambas medidas para el conjunto de modelos probit y logit, respectivamente. La primera columna de estas tablas recoge la réplica correspondiente, las columnas segunda y cuarta denotadas como D_1 y D_2 presentan la derivada en el punto medio, es decir, el instrumento tradicionalmente utilizado en la literatura para medir el cambio en probabilidad ante una variación unitaria de la respectiva variable explicativa (en este sentido, D_1 se asocia con x_1 y D_2 con x_2) y las otras dos columnas, DM_1 y DM_2, sintetizan la propuesta de medición efectuada basada en la media de las derivadas.

Centrándonos en la tabla 1 podemos concluir que de las 100 replicaciones sólo en cuatro de ellas (las numeradas como 7, 37, 49 y 79) la desigualdad obtenida es $D_K < DM_K \quad (K = 1,2)$; para el modelo logit (tabla 2), la anterior desigualdad no se presenta en ninguna ocasión. Por tanto, podemos establecer que de 100 casos —independientemente de que se trate de un modelo probit o logit— la mayoría de los mismos va a satisfacer que $D_K > DM_K$, lo que podemos entender como existencia de una pauta de comportamiento bastante sistemática entre ambas medidas.

Al objeto de poder aportar argumentos a favor de alguna de las medidas, parece razonable pensar en términos del criterio de «robustez». La robustez es una propiedad deseable de un buen estimador y determina, caso de cumplirse, que tal estimador sigue siendo razonablemente bueno ante pequeñas desviaciones del modelo.

En nuestro caso concreto, diremos que una medida es más robusta que la otra cuando ante pequeños cambios en la información muestral la primera altera su valor en menor medida que la segunda; en otros términos, tendremos más confianza en una medida cuyo comportamiento viene confirmado por toda la muestra que en otra que responda sólo a un subconjunto de individuos.

Para estudiar esta cuestión nos planteamos analizar, para cada muestra, la influencia que la eliminación de un individuo tiene sobre el valor de D_K o de DM_K en relación con el valor adoptado por la medida cuando hemos considerado $N = 100$. En concreto, se va eliminando cada vez un elemento distinto de la muestra, pasando a trabajar con muestras de tamaño 99, posteriormente, se compara el valor obtenido de D_K y DM_K, para cada muestra de tamaño 99, con el que denominamos valor de referencia que es el correspondiente a la muestra completa de tamaño 100.

Una forma adecuada de resumir las variaciones sufridas por D_K o DM_K, ante las pequeñas variaciones muestrales que hemos planteado, es atender a la dispersión o variación de las mismas calculada, para cada modelo, como:

(7.a) $$DD_i = \frac{\sum_{i=1}^{100} (D_{Ki} - VR_K)^2}{100}$$

$$(K = 1, 2)$$

(7.b) $$DDM_i = \frac{\sum_{i=1}^{100} \left(DM_{Ki} - VR'_K\right)^2}{100}$$

donde VR y VR' denotan los valores de referencia mencionados que indican, respectivamente, el valor de la derivada en el punto medio y la media de las derivadas para la muestra completa.

La aplicación de las expresiones dadas en (7.a) y (7.b) se efectua tanto para el enfoque probit como para el logit y las concreciones de las mismas aparecen, respectivamente, en las tablas 3 y 4 del Apéndice. Tales tablas presentan para cada réplica, numerada en la primera columna, los valores de (7.a) —columnas segunda y cuarta— para las dos variables explicativas que conforman la parte sistemática de cada modelo y, análogamente, la cuantificación de (7.b) aparece en las dos columnas restantes.

Tanto en el caso probit como en el logit se comprueba cómo, para todas las réplicas efectuadas, la medida D_K presenta una mayor dispersión —reflejada en el valor de DD_K— que DM_K.

Al objeto de resumir todo este conjunto de información, podemos calcular la media de estas dispersiones, es decir, hallar el promedio de cada una de las columnas que componen las tablas 3 y 4 y, en este sentido, definimos:

(8) $$PDD_K = \frac{\sum_{i=1}^{100} DD_i}{100} \qquad PDDM_K = \frac{\sum_{i=1}^{100} DDM_i}{100} \quad (K = 1, 2)$$

como los mencionados valores promedio, lo que nos permite obtener el conjunto de resultados siguientes:

	PDD_1	$PDDM_1$	PDD_2	$PDDM_2$
Probit	$0'000066838$	$0'000015144$	$0'000072187$	$0'000014424$
Logit	$0'000042917$	$0'000015422$	$0'000048772$	$0'000017606$

que reflejan, claramente, cómo para el caso probit, la dispersión de la medida basada en la derivada en el punto medio es en torno a $4'5-5$ veces la correspondiente a la propuesta alternativa, mientras que para el caso logit, esta relación resulta ser de $2'7$ veces[2].

Por tanto, atendiendo al criterio de robustez, podemos concluir que la propuesta defendida en este trabajo presenta una mayor estabilidad ante pequeños cambios en la información muestral y, en este sentido, resulta superior a la estrategia de actuación que habitualmente se viene utilizando para analizar las variaciones en la probabilidad de elección.

5. EJEMPLO: MODELOS PROBIT Y LOGIT APLICADOS A DATOS ELECTORALES

Con el fin de contrastar el comportamiento que, del ejercicio de simulación anteriormente desarrollado, se derivaba para las medidas D_K y DM_K, consideramos el siguiente caso real extraído de Aznar, García-Ferrer y Martín (1994). El estudio tiene como objetivo predecir la elección de los votantes sobre la construcción de un colegio local en Troy (Michigan) en el año 1973. La muestra está constituida por 95 personas y el conjunto de variables explicativas consideradas incluye las siguientes:

- X1: 1 si tiene uno o dos hijos en colegio público, 0 en otro caso.
- X2: 1 si tiene tres o cuatro hijos en colegio público, 0 en otro caso.
- X3: 1 si tiene cinco o más hijos en colegio público, 0 en otro caso.
- X4: 1 si tiene uno o más hijos en colegio privado, 0 en otro caso.
- X5: número de años viviendo en la comunidad de Troy.
- X6: 1 si el individuo es profesor (público o privado), 0 en otro caso.
- X7: logaritmo del ingreso familiar anual.
- X8: logaritmo del impuesto sobre bienes inmuebles pagado anualmente.

La variable de respuesta y_i se define como 1 si el individuo vota afirmativamente y 0 en otro caso.

En base a este conjunto de información se ajusta tanto un modelo logit como un modelo probit resultado:

[2] Resultados similares son obtenidos si razonamos, alternativamente, formando en base a las tablas 3 y 4 el cociente de dispersiones para cada réplica y cada x_i y resumiendo los 100 cocientes resultantes (para cada una de las variables explicativas) en su valor medio.

Variables	Coeficientes estimados	D_K	DM_K	DD_K	DDM_K	DD_K/DDM_K
		Modelo logit				
Término constante	$-5'20248$	–	–	–	–	–
X1	$0'58323$	–	–	–	–	–
X2	$1'12558$	–	–	–	–	–
X3	$0'52536$	–	–	–	–	–
X4	$-0'3415$	–	–	–	–	–
X5	$-0'026114$	$-0'0057961$	$-0'0050319$	$9'1104 \cdot 10^{-7}$	$6'31531 \cdot 10^{-7}$	$1'44259$
X6	$2'62739$	–	–	–	–	–
X7	$2'18742$	$0'48551$	$0'4215$	$0'00036434$	$0'00022363$	$1'62924$
X8	$-2'39464$	$-0'53151$	$-0'46143$	$0'00063372$	$0'00040738$	$1'55561$

Variables	Coeficientes estimados	D_K	DM_K	DD_K	DDM_K	DD_K/DDM_K
		Modelo probit				
Término constante	$-2'95916$	–	–	–	–	–
X1	$0'36807$	–	–	–	–	–
X2	$0'69117$	–	–	–	–	–
X3	$0'29522$	–	–	–	–	–
X4	$-0'21115$	–	–	–	–	–
X5	$-0'015761$	$-0'0057524$	$-0'0049942$	$7'21438 \cdot 10^{-7}$	$5'09011 \cdot 10^{-7}$	$1'41733$
X6	$1'58427$	–	–	–	–	–
X7	$1'31437$	$0'47971$	$0'41648$	$0'00035812$	$0'00021799$	$1'64279$
X8	$-1'46417$	$-0'53438$	$-0'46395$	$0'00058285$	$0'00037707$	$1'5457$

Notar que el conjunto de medidas analizadas en este trabajo sólo se ha calculado para las variables continuas, ya que únicamente eran este tipo de variables las que hemos considerado en el experimento de Monte Carlo.

La consideración de ambos modelos refleja que, en todos los casos, se satisface $|DM_K| < |D_K|$ lo cual coincide con la conclusión que, a este respecto, se desprendía del ejercicio de simulación. En cuanto a la dispersión de ambas medidas (DD_K y DDM_K) se mantiene la pauta de comportamiento esperada por cuanto $DD_K > DDM_K$ y ello se puede interpretar como evidencia a favor de la mayor «robustez» de esta última. Por último, con la consideración del cociente entre DD_K y DDM_K queremos proporcionar una cierta idea sobre la cuantía en que la dispersión de la medida habitualmente considerada supera a la propuesta efectuada; estos ratios también son coherentes con la conducta derivada en este sentido al trabajar con datos simulados.

6. CONCLUSIONES

El objetivo planteado en este trabajo ha consistido en efectuar algunas reflexiones sobre cómo analizar, en el marco de los modelos dicotómicos, la variación en la probabilidad de elección ante variaciones en una variable explicativa. Para ello, la discusión se ha centrado básicamente en dos cuestiones: en primer lugar, cuál debe ser el instrumento a utilizar y, posteriormente, cómo debe efectuarse la concreción de dicho instrumento.

En lo que respecta a la primera cuestión, se ha defendido que el cambio en probabilidad debe aproximarse a través de $\frac{\partial p_i}{\partial x_j}\Delta x_j$ para el caso de variables continuas, rechazando las críticas que algunos autores achacan a esta medida, en base a que el denominado problema de acotación no va a existir si se hace una interpretación correcta. Adicionalmente, en el supuesto de variables discretas el instrumento a utilizar es el análisis de sensibilidad.

En cuanto a la medición concreta del instrumento, en el apartado 3 se han tratado de sintetizar las diversas opciones existentes y se ha efectuado una valoración de cada una de ellas en términos del «uso eficiente de la información»; los puntos débiles detectados en las propuestas tradicionales se han tratado de subsanar planteando una concreción alternativa que atiende al valor medio de las derivadas, para el caso de variables continuas y, si se opera con variables ficticias, se defiende un análisis de sensibilidad en base a la media de cada submuestra concreta.

Establecido el marco de razonamiento alternativo, interesa profundizar en su comportamiento en relación a la opción comúnmente utilizada, basada en proceder a la cuantificación en la media de las variables. Para ello se ha planteado un ejercicio de simulación, que ha permitido verificar en primera instancia la existencia de una pauta de desenvolvimiento bastante sistemática entre ambas alternativas, por cuanto de los 100 casos analizados la casi totalidad de los mismos satisfacen $D_K > DM_K$ $(K = 1, 2)$, lo cual aunque, como ya se ha mencionado, no aporta ningún tipo de evidencia sobre la superioridad de una medida frente a la otra, sí que parece resultar más satisfactorio que el hecho de no haberse podido llegar a establecer ningún tipo de regularidad.

Planteado el tema en términos de criterios que permitan avalar la decantación hacia una de las estrategias, nos ha parecido oportuno razonar en base al criterio de robustez, el cual es ampliamente defendido en la literatura. El análisis de la misma para el marco de simulación planteado, nos permite concluir con bastante claridad que la propuesta alternativa expuesta en este trabajo, que entendemos constituye una vía coherente de determinar el comportamiento medio de la muestra, es más robusta que la habitualmente utilizada.

Las pautas de comportamiento derivadas del ejercicio de simulación se ven refrendadas en el caso real planteado como ejemplo. En definitiva, el cumplimiento de la propiedad de «robustez» por parte de la estrategia de actuación propuesta, es un resultado alentador para proseguir con el estudio de la misma y profundizar en otros aspectos de su comportamiento.

APÉNDICE

Tabla 1. *Modelo probit*

RÉPLICA	D_1	DM_1	D_2	DM_2
1	0.29938	0.20887	0.34193	0.23856
2	0.25206	0.19267	0.25817	0.19734
3	0.26402	0.21164	0.16769	0.13442
4	0.20228	0.17255	0.18783	0.16022
5	0.20141	0.15969	0.23990	0.19021
6	0.28364	0.22509	0.26436	0.20979
7	0.26788	0.26918	0.18100	0.18188
8	0.38533	0.27581	0.27001	0.19327
9	0.32265	0.21458	0.38220	0.25418
10	0.34428	0.23993	0.33918	0.23637
11	0.33830	0.22307	0.37752	0.24893
12	0.25637	0.20564	0.19293	0.15475
13	0.37863	0.24019	0.42532	0.26980
14	0.19163	0.16002	0.21051	0.17579
15	0.33915	0.24698	0.29297	0.21336
16	0.31129	0.21825	0.33449	0.23451
17	0.24867	0.21968	0.22481	0.19861
18	0.21752	0.17876	0.22163	0.18214
19	0.27778	0.23311	0.23363	0.19606
20	0.23542	0.17427	0.31659	0.23435
21	0.26785	0.18142	0.37703	0.25538
22	0.13649	0.11501	0.24746	0.20851
23	0.44524	0.26938	0.39697	0.24018
24	0.19847	0.19253	0.20090	0.19488
25	0.24605	0.22545	0.19378	0.17755
26	0.19499	0.16705	0.19152	0.16407
27	0.22320	0.17766	0.24625	0.19600
28	0.29108	0.20306	0.33456	0.23339
29	0.14784	0.13152	0.24129	0.21466
30	0.28622	0.20994	0.24379	0.17882

Tabla 1. (cont.) *Modelo probit*

RÉPLICA	D_1	DM_1	D_2	DM_2
31	0.31340	0.21262	0.33465	0.22703
32	0.29649	0.21712	0.28364	0.20771
33	0.096313	0.087231	0.22809	0.20658
34	0.27648	0.20863	0.24601	0.18563
35	0.26712	0.21922	0.24842	0.20387
36	0.28245	0.24792	0.19531	0.17143
37	0.11215	0.13527	0.18557	0.22382
38	0.39264	0.25873	0.31547	0.20788
39	0.34028	0.25137	0.26449	0.19539
40	0.30951	0.22780	0.23517	0.17308
41	0.29030	0.21551	0.25986	0.19291
42	0.35914	0.23747	0.37139	0.24557
43	0.36605	0.25654	0.25050	0.17556
44	0.28139	0.23291	0.24061	0.19916
45	0.29130	0.20350	0.38184	0.26675
46	0.28067	0.20847	0.29261	0.21734
47	0.24974	0.18072	0.31269	0.22628
48	0.27215	0.21252	0.30426	0.23760
49	0.17140	0.17941	0.15079	0.15784
50	0.22777	0.15715	0.36917	0.25471
51	0.27075	0.18425	0.36210	0.24642
52	0.25616	0.16060	0.44682	0.28013
53	0.33707	0.24286	0.32702	0.23561
54	0.30920	0.21593	0.30580	0.21356
55	0.23494	0.18705	0.23443	0.18664
56	0.26018	0.18247	0.33154	0.23251
57	0.21808	0.19425	0.20238	0.18026
58	0.29230	0.23741	0.25878	0.21019
59	0.28221	0.21273	0.24903	0.18772
60	0.28184	0.20389	0.28269	0.20450
61	0.22339	0.17059	0.29902	0.22834

Tabla 1. (cont.) *Modelo probit*

RÉPLICA	D_1	DM_1	D_2	DM_2
62	0.22710	0.18496	0.23454	0.19102
63	0.30099	0.21268	0.31603	0.22330
64	0.34976	0.24539	0.26866	0.18849
65	0.30349	0.21257	0.30550	0.21397
66	0.33462	0.21400	0.39545	0.25290
67	0.20487	0.17068	0.17282	0.14397
68	0.29225	0.21871	0.23739	0.17765
69	0.42138	0.28620	0.25751	0.17490
70	0.19093	0.18225	0.18985	0.18122
71	0.25454	0.19239	0.25316	0.19134
72	0.14537	0.14845	0.19110	0.19515
73	0.18628	0.16214	0.21552	0.18759
74	0.33545	0.23216	0.37501	0.25954
75	0.20817	0.16935	0.20457	0.16642
76	0.23933	0.15604	0.42059	0.27422
77	0.34932	0.24054	0.33695	0.23203
78	0.27095	0.20161	0.25842	0.19228
79	0.24827	0.17884	0.34808	0.25074
80	0.20303	0.15594	0.28174	0.21640
81	0.26513	0.17136	0.42101	0.27210
82	0.38170	0.26500	0.25691	0.17836
83	0.30005	0.20460	0.34534	0.23548
84	0.25534	0.21587	0.27301	0.23081
85	0.26048	0.23819	0.25182	0.23028
86	0.32725	0.23696	0.23745	0.17194
87	0.22519	0.17173	0.28059	0.21398
88	0.24001	0.18093	0.29534	0.22264
89	0.21894	0.17112	0.27758	0.21696
90	0.28828	0.19819	0.33403	0.22964
91	0.16827	0.14208	0.22265	0.18799
92	0.36031	0.22679	0.39272	0.24719

Tabla 1. (cont.) *Modelo probit*

RÉPLICA	D_1	DM_1	D_2	DM_2
93	0.21183	0.18061	0.19200	0.16370
94	0.22954	0.16713	0.31868	0.23203
95	0.24285	0.19125	0.22323	0.17581
96	0.21608	0.16085	0.29430	0.21908
97	0.36163	0.25448	0.31262	0.21999
98	0.15177	0.12168	0.26914	0.21578
99	0.24362	0.21548	0.18457	0.16325
100	0.30931	0.22273	0.31963	0.23016

Tabla 2. *Modelo logit*

RÉPLICA	D_1	DM_1	D_2	DM_2
1	0.12643	0.11286	0.13344	0.11912
2	0.29087	0.20675	0.27438	0.19503
3	0.23200	0.17188	0.27366	0.20275
4	0.23012	0.18725	0.11930	0.097080
5	0.21939	0.18114	0.23800	0.19651
6	0.34125	0.23440	0.28322	0.19453
7	0.23245	0.18256	0.20433	0.16048
8	0.25008	0.18073	0.28744	0.20773
9	0.21919	0.18021	0.15098	0.12413
10	0.21716	0.17558	0.19102	0.15445
11	0.19225	0.15137	0.23253	0.18308
12	0.20060	0.15690	0.23138	0.18098
13	0.21578	0.18116	0.13520	0.11351
14	0.22882	0.17747	0.25494	0.19774
15	0.11067	0.097809	0.16274	0.14383
16	0.24914	0.19021	0.23022	0.17577
17	0.19494	0.16411	0.18147	0.15277
18	0.10331	0.089947	0.17714	0.15422
19	0.11263	0.093074	0.23212	0.19182
20	0.21868	0.15820	0.30157	0.21816
21	0.30306	0.21066	0.28941	0.20117
22	0.16894	0.13928	0.18911	0.15590
23	0.25055	0.20153	0.13847	0.11138
24	0.16265	0.13441	0.19986	0.16515
25	0.25603	0.17697	0.33304	0.23020
26	0.27135	0.20785	0.18520	0.14186
27	0.22403	0.18509	0.14285	0.11802
28	0.16758	0.13611	0.23552	0.19129
29	0.16596	0.14228	0.18637	0.15977
30	0.23107	0.19726	0.069065	0.058961

Tabla 2. (cont.) *Modelo logit*

RÉPLICA	D_1	DM_1	D_2	DM_2
31	0.16654	0.14427	0.13193	0.11428
32	0.13024	0.10862	0.20762	0.17315
33	0.37474	0.23885	0.40485	0.25803
34	0.19035	0.15690	0.17847	0.14710
35	0.14242	0.12405	0.18896	0.16457
36	0.15076	0.11244	0.30437	0.22700
37	0.22316	0.17115	0.23151	0.17756
38	0.29836	0.21477	0.21728	0.15641
39	0.19806	0.15651	0.27506	0.21736
40	0.17457	0.15082	0.12499	0.10799
41	0.19443	0.16416	0.14038	0.11853
42	0.24458	0.18184	0.27108	0.20155
43	0.22461	0.17998	0.18147	0.14542
44	0.17199	0.13725	0.25424	0.20289
45	0.086481	0.077286	0.16440	0.14692
46	0.25485	0.20695	0.11735	0.095294
47	0.25133	0.18906	0.22534	0.16951
48	0.19168	0.15864	0.17106	0.14158
48	0.15583	0.13349	0.16681	0.14290
50	0.23465	0.18464	0.17562	0.13819
51	0.15666	0.13493	0.16323	0.14059
52	0.28036	0.20776	0.21292	0.15778
53	0.25333	0.19686	0.15819	0.12292
54	0.22715	0.18338	0.19457	0.15707
55	0.36606	0.24552	0.27974	0.18763
56	0.17666	0.14286	0.21868	0.17683
57	0.30034	0.20461	0.31595	0.21524
58	0.15569	0.12452	0.24297	0.19433
59	0.28016	0.19414	0.33138	0.22963
60	0.24730	0.17429	0.31898	0.22480
61	0.16627	0.14711	0.17176	0.15196

Tabla 2. (cont.) *Modelo logit*

RÉPLICA	D_1	DM_1	D_2	DM_2
62	0.14556	0.12728	0.13708	0.11986
63	0.16860	0.13843	0.21021	0.17259
64	0.19107	0.15952	0.16905	0.14113
65	0.39247	0.26247	0.10689	0.071482
66	0.15087	0.13865	0.095620	0.087875
67	0.19403	0.15919	0.20832	0.17092
68	0.24857	0.20178	0.13518	0.10973
69	0.22022	0.16320	0.27835	0.20627
70	0.18310	0.15186	0.16338	0.13551
71	0.17588	0.14123	0.21507	0.17270
72	0.22089	0.17112	0.24111	0.18678
73	0.24333	0.18565	0.21380	0.16312
74	0.23589	0.18036	0.24787	0.18951
75	0.13953	0.12078	0.16696	0.14452
76	0.19747	0.14911	0.27977	0.21126
77	0.24395	0.19153	0.18290	0.14359
78	0.22249	0.17105	0.23159	0.17804
79	0.22160	0.17565	0.17965	0.14240
80	0.14873	0.12870	0.15971	0.13821
81	0.18939	0.15049	0.22113	0.17571
82	0.27033	0.21004	0.16652	0.12938
83	0.17161	0.14605	0.15652	0.13321
84	0.24429	0.18261	0.25272	0.18891
85	0.18912	0.13703	0.34571	0.25049
86	0.27394	0.17280	0.42165	0.26598
87	0.14242	0.12101	0.17518	0.14885
88	0.27106	0.19826	0.27452	0.20079
89	0.17779	0.15893	0.10390	0.092873
90	0.21918	0.17953	0.13911	0.11394
91	0.14962	0.13435	0.081254	0.072964
92	0.32357	0.21130	0.37521	0.24501

Tabla 2. (cont.) *Modelo logit*

RÉPLICA	D_1	DM_1	D_2	DM_2
93	0.20552	0.16392	0.20170	0.16087
94	0.21319	0.16788	0.20840	0.16411
95	0.24501	0.17748	0.29683	0.21502
96	0.28377	0.18461	0.38142	0.24813
97	0.20760	0.16430	0.21772	0.17231
98	0.14742	0.13027	0.12265	0.10838
99	0.23483	0.18213	0.19721	0.15295
100	0.15070	0.13492	0.11110	0.099465

Tabla 3. *Modelo probit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
1	0.6623200E-04	0.1496600E-04	0.9964000E-04	0.1868400E-04
2	0.5145000E-04	0.1785500E-04	0.5155200E-04	0.1561800E-04
3	0.3810700E-04	0.1158600E-04	0.2999600E-04	0.1424000E-04
4	0.3540700E-04	0.1661700E-04	0.3412500E-04	0.1663700E-04
5	0.4137200E-04	0.1840100E-04	0.4152400E-04	0.1411200E-04
6	0.7812200E-04	0.1663600E-04	0.7927500E-04	0.1533200E-04
7	0.1941900E-03	0.1675100E-04	0.9422800E-04	0.1220300E-04
8	0.1483000E-03	0.1498900E-04	0.8937400E-04	0.1867900E-04
9	0.7918500E-04	0.1491800E-04	0.1076500E-03	0.1509300E-04
10	0.1015600E-03	0.1528500E-04	0.1016500E-03	0.1499000E-04
11	0.8056800E-04	0.1491600E-04	0.1063100E-03	0.1441600E-04
12	0.4438700E-04	0.2046300E-04	0.3136700E-04	0.1180000E-04
13	0.1240500E-03	0.1441400E-04	0.1590800E-03	0.1172800E-04
14	0.2436000E-04	0.9411000E-05	0.3266500E-04	0.1183900E-04
15	0.9469000E-04	0.1405700E-04	0.8387400E-04	0.1413600E-04
16	0.7340500E-04	0.1370300E-04	0.8433300E-04	0.1398800E-04
17	0.7819400E-04	0.1284500E-04	0.7175900E-04	0.1754900E-04
18	0.3657600E-04	0.1308700E-04	0.4141600E-04	0.1412600E-04
19	0.9208500E-04	0.1623300E-04	0.7832800E-04	0.1746400E-04
20	0.4801400E-04	0.1387800E-04	0.8179400E-04	0.1611400E-04
21	0.5698300E-04	0.1569600E-04	0.8038100E-04	0.1257700E-04
22	0.2142900E-04	0.1220200E-04	0.4716300E-04	0.1233100E-04
23	0.1070200E-03	0.1396500E-04	0.1089800E-03	0.1645900E-04
24	0.5854300E-04	0.1913000E-04	0.6019100E-04	0.1855400E-04
25	0.1001800E-03	0.1590100E-04	0.6919700E-04	0.1584700E-04
26	0.2548600E-04	0.1083200E-04	0.2821200E-04	0.9785700E-05
27	0.4230300E-04	0.1586100E-04	0.4408400E-04	0.1298700E-04
28	0.4563500E-04	0.9041200E-05	0.7309100E-04	0.1470800E-04
29	0.2912100E-04	0.8938400E-05	0.6485900E-04	0.9873100E-05
30	0.3483300E-04	0.9293600E-05	0.4100200E-04	0.1379400E-04

Tabla 3. (cont.) *Modelo probit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
31	0.7699500E-04	0.1632300E-04	0.9149700E-04	0.1814600E-04
32	0.5513500E-04	0.1266000E-04	0.5947500E-04	0.1272700E-04
33	0.2274500E-04	0.1225900E-04	0.4637200E-04	0.1267100E-04
34	0.4111600E-04	0.1025500E-04	0.5019800E-04	0.1600300E-04
35	0.7461000E-04	0.1480100E-04	0.6726900E-04	0.1242700E-04
36	0.1045300E-03	0.2178900E-04	0.6044800E-04	0.2351200E-04
37	0.4910400E-04	0.1139200E-04	0.1356700E-03	0.1349800E-04
38	0.8339800E-04	0.1164900E-04	0.7375900E-04	0.1452700E-04
39	0.9536600E-04	0.1515800E-04	0.6193600E-04	0.1318700E-04
40	0.4688900E-04	0.1086300E-04	0.3594100E-04	0.1050800E-04
41	0.4974100E-04	0.1214900E-04	0.4647800E-04	0.1332500E-04
42	0.1041300E-03	0.1766100E-04	0.1143200E-03	0.1556300E-04
43	0.7336200E-04	0.1131700E-04	0.4870400E-04	0.1338500E-04
44	0.9079900E-04	0.1264600E-04	0.7108800E-04	0.1318600E-04
45	0.9418500E-04	0.1686800E-04	0.1587700E-03	0.1304000E-04
46	0.5438100E-04	0.1274900E-04	0.6397100E-04	0.1203700E-04
47	0.5484200E-04	0.1761700E-04	0.5627100E-04	0.1108400E-04
48	0.9550100E-04	0.1666000E-04	0.1197400E-03	0.1768500E-04
49	0.4465900E-04	0.1289300E-04	0.3840000E-04	0.1133900E-04
50	0.4802500E-04	0.1675600E-04	0.7789200E-04	0.1123500E-04
51	0.4787300E-04	0.1160900E-04	0.9292300E-04	0.1648600E-04
52	0.6761400E-04	0.1639800E-04	0.9443300E-04	0.1022000E-04
53	0.1148600E-03	0.1269500E-04	0.1110500E-03	0.1740300E-04
54	0.5111600E-04	0.1240800E-04	0.4993200E-04	0.1099100E-04
55	0.3782100E-04	0.1226600E-04	0.4430200E-04	0.1521000E-04
56	0.4474300E-04	0.1155900E-04	0.6105500E-04	0.1207500E-04
57	0.5030100E-04	0.1434000E-04	0.5402500E-04	0.1439600E-04
58	0.1003200E-03	0.1780600E-04	0.8739200E-04	0.1463700E-04
59	0.5476700E-04	0.1483600E-04	0.5350100E-04	0.1677300E-04
60	0.4867400E-04	0.1433900E-04	0.4175800E-04	0.9769500E-05
61	0.4921000E-04	0.1678600E-04	0.7251400E-04	0.1589300E-04

Tabla 3. (cont.) *Modelo probit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
62	0.3909300E-04	0.1348100E-04	0.4485400E-04	0.1005300E-04
63	0.5543200E-04	0.1204400E-04	0.6086800E-04	0.1168200E-04
64	0.6414100E-04	0.1153900E-04	0.5335000E-04	0.1359700E-04
65	0.4996100E-04	0.1235500E-04	0.5755300E-04	0.1384800E-04
66	0.7529100E-04	0.1287700E-04	0.1303000E-03	0.2008500E-04
67	0.2467600E-04	0.1014500E-04	0.2444800E-04	0.1125400E-04
68	0.4545100E-04	0.1190400E-04	0.4765400E-04	0.1613000E-04
69	0.1510200E-03	0.1625600E-04	0.7534000E-04	0.1532100E-04
70	0.4562300E-04	0.1339100E-04	0.5103200E-04	0.1431500E-04
71	0.5684400E-04	0.1953500E-04	0.5523600E-04	0.1801800E-04
72	0.4474600E-04	0.1095500E-04	0.6879100E-04	0.1102900E-04
73	0.3381200E-04	0.1464100E-04	0.4533900E-04	0.1365900E-04
74	0.1223000E-03	0.1897600E-04	0.1568100E-03	0.1312300E-04
75	0.3535300E-04	0.1576600E-04	0.3626800E-04	0.1477400E-04
76	0.7060100E-04	0.1869100E-04	0.1177800E-03	0.1338600E-04
77	0.9684800E-04	0.1597900E-04	0.1009900E-03	0.1598300E-04
78	0.4084800E-04	0.1153600E-04	0.4896200E-04	0.1490900E-04
79	0.5668100E-04	0.1254700E-04	0.9856500E-04	0.1166300E-04
80	0.2917300E-04	0.9739300E-05	0.5411200E-04	0.1379400E-04
81	0.7697900E-04	0.2554500E-04	0.1042400E-03	0.1268800E-04
82	0.8923200E-04	0.1252800E-04	0.5220600E-04	0.1082700E-04
83	0.5253100E-04	0.1185900E-04	0.7515800E-04	0.1452000E-04
84	0.1869000E-03	0.5635800E-04	0.2458700E-03	0.3370500E-04
85	0.1514100E-03	0.1495200E-04	0.1394200E-03	0.1250600E-04
86	0.5912600E-04	0.1295400E-04	0.4838000E-04	0.1475800E-04
87	0.3423700E-04	0.1064600E-04	0.5029500E-04	0.1224500E-04
88	0.4477500E-04	0.1564300E-04	0.5608300E-04	0.1266800E-04
89	0.4787800E-04	0.1804200E-04	0.5366300E-04	0.1014000E-04
90	0.7247300E-04	0.1881600E-04	0.6756500E-04	0.1292900E-04
91	0.2296300E-04	0.9895300E-05	0.3559500E-04	0.1216000E-04
92	0.1032600E-03	0.1887900E-04	0.9917200E-04	0.1425800E-04

Tabla 3. (cont.) *Modelo probit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
93	0.3953400E-04	0.1861400E-04	0.3478300E-04	0.1583000E-04
94	0.1365400E-03	0.5495300E-04	0.1186100E-03	0.3486100E-04
95	0.3152000E-04	0.9275400E-05	0.3727200E-04	0.1333100E-04
96	0.4714500E-04	0.1692500E-04	0.5345900E-04	0.1306600E-04
97	0.1101200E-03	0.1425700E-04	0.8603700E-04	0.1398300E-04
98	0.2135200E-04	0.8661000E-05	0.4578500E-04	0.1141000E-04
99	0.6002100E-04	0.1655200E-04	0.4724300E-04	0.1924500E-04
100	0.7932300E-04	0.1534300E-04	0.8141000E-04	0.1208500E-04

Tabla 4. *Modelo logit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
1	0.3317700E-04	0.2070100E-04	0.3496000E-04	0.2150500E-04
2	0.8078600E-04	0.2143000E-04	0.6342400E-04	0.1654700E-04
3	0.5892300E-04	0.1933900E-04	0.5916000E-04	0.1555200E-04
4	0.3763900E-04	0.1281600E-04	0.4071800E-04	0.2259200E-04
5	0.3856500E-04	0.1240400E-04	0.5192800E-04	0.1447600E-04
6	0.7488700E-04	0.1292500E-04	0.7963400E-04	0.2006000E-04
7	0.3740600E-04	0.1374600E-04	0.4471500E-04	0.1886100E-04
8	0.4728300E-04	0.1480200E-04	0.4589500E-04	0.1099400E-04
9	0.3794700E-04	0.1522500E-04	0.3455500E-04	0.1777500E-04
10	0.4995000E-04	0.2064500E-04	0.4358800E-04	0.1855100E-04
11	0.4148700E-04	0.1792000E-04	0.4302000E-04	0.1542300E-04
12	0.3215500E-04	0.1094600E-04	0.6320900E-04	0.2095200E-04
13	0.3148100E-04	0.1315600E-04	0.2298500E-04	0.1264800E-04
14	0.3792400E-04	0.1377200E-04	0.4708900E-04	0.1320400E-04
15	0.2209500E-04	0.1381000E-04	0.4808500E-04	0.2569900E-04
16	0.8289800E-04	0.2734500E-04	0.8393600E-04	0.2888100E-04
17	0.3967600E-04	0.2008500E-04	0.4451200E-04	0.2269400E-04
18	0.2349800E-04	0.1539800E-04	0.4638400E-04	0.2405900E-04
19	0.2370300E-04	0.1424000E-04	0.4043900E-04	0.1526900E-04
20	0.5901500E-04	0.2118500E-04	0.5437300E-04	0.1284200E-04
21	0.5345100E-04	0.1216700E-04	0.5618400E-04	0.1370100E-04
22	0.3060500E-04	0.1411900E-04	0.4331800E-04	0.1912600E-04
23	0.3588100E-04	0.1150700E-04	0.2377200E-04	0.1204500E-04
24	0.2910900E-04	0.1357200E-04	0.5079100E-04	0.2209600E-04
25	0.5472000E-04	0.1483000E-04	0.5808100E-04	0.1031300E-04
26	0.6337600E-04	0.1765400E-04	0.7972900E-04	0.3306900E-04
27	0.5039100E-04	0.2112000E-04	0.2938200E-04	0.1584600E-04
28	0.3070400E-04	0.1451300E-04	0.3763700E-04	0.1233600E-04
29	0.2505200E-04	0.1332000E-04	0.2870600E-04	0.1287600E-04
30	0.3330200E-04	0.1244600E-04	0.2488700E-04	0.1723300E-04

Tabla 4. (cont.) *Modelo logit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
31	0.2682100E-04	0.1445200E-04	0.2755100E-04	0.1693700E-04
32	0.3327300E-04	0.1870700E-04	0.4120200E-04	0.1732400E-04
33	0.1200500E-03	0.1598000E-04	0.1381200E-03	0.1409000E-04
34	0.2721500E-04	0.1186700E-04	0.3334800E-04	0.1574400E-04
35	0.2193700E-04	0.1280400E-04	0.2753100E-04	0.1217600E-04
36	0.3496800E-04	0.1591700E-04	0.6245300E-04	0.1479100E-04
37	0.5297400E-04	0.1942700E-04	0.5065600E-04	0.1782900E-04
38	0.4945400E-04	0.1126300E-04	0.4975500E-04	0.1813800E-04
39	0.3336200E-04	0.1021000E-04	0.5704000E-04	0.1232400E-04
40	0.2464000E-04	0.1268600E-04	0.2270600E-04	0.1417500E-04
41	0.3241400E-04	0.1477700E-04	0.2884300E-04	0.1633300E-04
42	0.5256100E-04	0.1670100E-04	0.6070300E-04	0.1620900E-04
43	0.2704900E-04	0.9646000E-05	0.3194100E-04	0.1493100E-04
44	0.2948800E-04	0.1149900E-04	0.5012900E-04	0.1396700E-04
45	0.2155500E-04	0.1493800E-04	0.4202600E-04	0.2375900E-04
46	0.4048600E-04	0.1276600E-04	0.3079200E-04	0.1683300E-04
47	0.4051700E-04	0.1290600E-04	0.4083400E-04	0.1518800E-04
48	0.3838300E-04	0.1789600E-04	0.3162700E-04	0.1577400E-04
49	0.3226700E-04	0.1826800E-04	0.3582000E-04	0.1918000E-04
50	0.5106800E-04	0.1608700E-04	0.4784800E-04	0.2027000E-04
51	0.3697300E-04	0.2103400E-04	0.2799800E-04	0.1457800E-04
52	0.4910500E-04	0.1299800E-04	0.4871700E-04	0.1748100E-04
53	0.4162400E-04	0.1270200E-04	0.4134400E-04	0.2016600E-04
54	0.2884300E-04	0.9620900E-05	0.3092600E-04	0.1313900E-04
55	0.8042600E-04	0.1485300E-04	0.6569100E-04	0.1702700E-04
56	0.3122800E-04	0.1331400E-04	0.6717200E-04	0.2711100E-04
57	0.7690300E-04	0.1625000E-04	0.8501300E-04	0.1781100E-04
58	0.2917400E-04	0.1411900E-04	0.4472800E-04	0.1489100E-04
59	0.5177800E-04	0.1325800E-04	0.6892200E-04	0.1283100E-04
60	0.6409000E-04	0.1914100E-04	0.6601600E-04	0.1329400E-04
61	0.2115400E-04	0.1102400E-04	0.2234300E-04	0.1074900E-04

Tabla 4. (cont.) *Modelo logit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
62	0.3012200E-04	0.1712200E-04	0.3738400E-04	0.2253200E-04
63	0.3010400E-04	0.1474600E-04	0.3601700E-04	0.1466100E-04
64	0.4268700E-04	0.2026800E-04	0.4480300E-04	0.2282100E-04
65	0.8219200E-04	0.1024400E-04	0.9083400E-04	0.3465700E-04
66	0.1723700E-04	0.1049200E-04	0.2503100E-04	0.1876000E-04
67	0.3603700E-04	0.1772200E-04	0.3667400E-04	0.1466200E-04
68	0.4421800E-04	0.1490600E-04	0.2891000E-04	0.1618800E-04
69	0.4355600E-04	0.1518900E-04	0.5501200E-04	0.1507800E-04
70	0.3499600E-04	0.1623700E-04	0.3002300E-04	0.1545200E-04
71	0.3874400E-04	0.1746600E-04	0.5240200E-04	0.2061300E-04
72	0.3752800E-04	0.1429600E-04	0.4295200E-04	0.1385200E-04
73	0.8357200E-04	0.2558700E-04	0.8374500E-04	0.2913900E-04
74	0.5480200E-04	0.1732100E-04	0.7026700E-04	0.2210900E-04
75	0.2663500E-04	0.1612800E-04	0.3014600E-04	0.1637100E-04
76	0.4005300E-04	0.1480400E-04	0.6144900E-04	0.1644600E-04
77	0.5245000E-04	0.1888800E-04	0.3535500E-04	0.1605100E-04
78	0.7161300E-04	0.2566300E-04	0.5355700E-04	0.1711400E-04
79	0.3599200E-04	0.1378400E-04	0.4627600E-04	0.2283400E-04
80	0.2331400E-04	0.1314800E-04	0.4176900E-04	0.2351400E-04
81	0.3516900E-04	0.1488000E-04	0.4019900E-04	0.1483500E-04
82	0.5167600E-04	0.1540100E-04	0.3793500E-04	0.1700000E-04
83	0.2970700E-04	0.1497600E-04	0.3182500E-04	0.1686700E-04
84	0.3396600E-04	0.1031800E-04	0.4965400E-04	0.1658800E-04
85	0.5126700E-04	0.1706500E-04	0.8714900E-04	0.1265400E-04
86	0.7312300E-04	0.1619000E-04	0.1056100E-03	0.1268900E-04
87	0.2685200E-04	0.1430400E-04	0.3877500E-04	0.1893200E-04
88	0.4812800E-04	0.1438600E-04	0.5483700E-04	0.1454900E-04
89	0.3042100E-04	0.1772700E-04	0.2293900E-04	0.1574500E-04
90	0.3534500E-04	0.1361100E-04	0.2732500E-04	0.1479600E-04
91	0.2465000E-04	0.1428800E-04	0.3180100E-04	0.2361900E-04
92	0.8940500E-04	0.1815200E-04	0.1176100E-03	0.1803600E-04

Tabla 4. (cont.) *Modelo logit. Medida de dispersión*

RÉPLICA	DD_1	DDM_1	DD_2	DDM_2
93	0.5491700E-04	0.1997900E-04	0.6680000E-04	0.2605700E-04
94	0.4583000E-04	0.1722700E-04	0.5007900E-04	0.1974600E-04
95	0.4535900E-04	0.1546300E-04	0.5517400E-04	0.1319300E-04
96	0.6988900E-04	0.1438800E-04	0.1031800E-03	0.1550300E-04
97	0.2639500E-04	0.9943600E-05	0.4484800E-04	0.1762500E-04
98	0.2805400E-04	0.1601300E-04	0.2770300E-04	0.1759700E-04
99	0.3763400E-04	0.1187700E-04	0.4539300E-04	0.1835800E-04
100	0.2920700E-04	0.1775800E-04	0.3084500E-04	0.2103600E-04

REFERENCIAS

[1] **Aldrich, J.N.** and **F.D. Nelson** (1984). *Linear Probability, Logit and Probit Models*. (Sage, California).

[2] **Amemiya, T.** (1981). «Qualitative response models: A survey», *Journal of Economic Literature*, **19**, 1483–1536.

[3] **Aznar, A., A. García-Ferrer** y **A. Martín** (1994). *Ejercicios de Econometría*. Volumen II (Pirámide, Madrid).

[4] **Fisher, T.C.G.** (1991). «An empirical study of the adverse selection model of strikes», *Canadian Journal of Economics*, **XXIV**, 499–516.

[5] **Gourieroux, C., A. Monfort, E. Renault** and **A. Trognon** (1987). «Simulated residuals», *Journal of Econometrics*, **34**, 201–252.

[6] **Greene, W.H.** (1993). *Econometric Analysis*. Second Edition (MacMillan, New York).

[7] **Greene, L.** and **T. Seaks** (1992). «An analysis of the probability of default on federally guaranteed student loans», *Review of Economics and Statistics*, **74**, 404–411.

[8] **Griffiths, W.E., R. Carter-Hill** and **P.J. Pope** (1987). «Small sample properties of probit model estimators», *Journal of the American Statistical Association*, **82**, 929–937.

[9] **Hill, J.W.** y **R.W. Ingram** (1989). «Selection of GAAP of RAP in the Savings and Loan Industry», *The Accounting Review*, **LXIV, 4**, 667–679.

[10] **Jhonson, T.** (1987). *The analysis of qualitative and limited responses*, in: W.E. Becker and W.B. Walstad, eds., Econometric modeling in economic education research (Kluwer-Nijhoff, Boston).

[11] **LeClere, M.J.** (1992). «The interpretation of coefficients in models with qualitative dependent variables», *Decision Science Journal*, **23**, 770–776.

[12] **Maddala, G.S.** (1983). *Limited dependent and qualitative variables in Econometrics*. (Cambridge University, New York).

[13] **Maddala, G.S.** (1992). *Introduction to Econometrics*. Second Edition, (Mac Millan, New York).

[14] **McFadden, D.** (1976). «Quantal Choice analysis: A survey», *Annals of Economic and Social Measurement*, **5**, 363–390.

[15] **Pagan, A.** and **F. Vella** (1989). «Diagnostic tests for model based on individual data: A survey», *Journal of Applied Econometric*, **4**, 19–59.

[16] **Petersen, T.** (1985). «A comment of presenting results from logit and probit models», *American Sociological Review*, **50**, 130–131.

[17] **Train, K.** (1986). *Qualitative choice analysis*. (MIT Press, Massachusetts).

[18] **Windmeijer, F.A.G.** (1995). «Goodness-of-fit measures in binary choice models», *Economic Reviews*, **14**, 101–116.

ENGLISH SUMMARY

REFLECTIONS ON THE PROBABILITY CHANGE MEASUREMENT STRATEGY IN BINARY CHOICE MODELS

M.T. APARICIO*
I. VILLANÚA*
Universidad de Zaragoza⋆

This paper focuses on the evaluation of the measure which, in the context of Binary Choice Models, is required in order to represent the probability variation in the case of changes in the independent variables. The most common option to quantify this measure consists in evaluating it in the vector of mean values of the different explanatory variables. Our alternative strategy considers a more efficient use of sample information, based on emphasizing the average behaviour of the sample, rather than the behaviour of an «*average individual*»*. Finally, a simulation exercise is presented to illustrate that the proposed approach is more robust than the traditional option.*

Keywords: Discrete choice, probability variation, robustness.

AMS Classification: 62J02, 65C05

---

* Los autores desean agradecer los comentarios de Antonio Aznar, Carmen García-Olaverri y de un evaluador anónimo, así como la financiación de DGICYT PB94-0602.

⋆ Departamento de Análisis Económico. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza. Gran Vía, 2. 50005 Zaragoza.

– Received July 1996.
– Accepted January 1998.

In recent years the so-called discrete choice models (DCM) have enjoyed considerable success due to the greater emphasis that has been placed on disaggregated models in econometric research. This type of model attempts to obtain the probability that a decision-making unit choose a given alternative from among a set of options, when the response regarding the choice made (y) and a set of variables (x) which represent characteristics of the available choices and of the decision-making unit are know.

The simplest models in this category are the dichotomous or binary DCMs, which can be represented generally as follows:

$$p_i = F(x_i' \beta) \tag{1}$$

and upon which this paper will focus.

In spite of the large number of empirical studies which use these models, their validation stage usually receives little attention. It is normal to present the estimated coefficients and their corresponding $t-$ratios, as well as one of the multiple scale measures of goodness of fit proposed for this kind of approach, the LR test relative to the joint significance of the model and the analysis of the effect the variations of the different socio-economic (explanatory) variables have on the probability of choice. The aim of this paper is to study only this last mentioned aspect.

In our context (non-linear models), this effect is determined, for continuous variables, by the expression:

$$(\partial p_i / \partial x_j)\, \Delta x_j \tag{2}$$

and, for discrete variables, by:

$$\frac{\Delta p_i}{\Delta x_j} = \frac{P(y_i = 1 | x_j = 1) - P(y_i = 1 | x_j = 0)}{1 - 0} \tag{3}$$

In either of these two situations, the specification of the probability variation requires the calculation of $x_i' \beta$. The parameter vector will have been estimated previously, but we need to decide on the value to be adopted for the set of independent variables. In this respect, we cannot speak of a unique strategy; rather, the literature contains various approaches. The most frequent option is to evaluate in the vector of the average mean values of the x's. In our opinion, this practice has some defects. First, in the case of dummy variables, the mean value is not very coherent with the meaning given to these variables. Secondly, the use of the average of the sample can lead into problems of lack of representativeness if the values of a given variable are very disperse.

Another option is to calculate $(\partial p_i / \partial x_j)$, in what are generically referred to as the "most significant"values of the explanatory variables. This practice is somewhat ambiguous, since the meaning of the term «most significant» is not explicitly defined.

In our view, the relevant question is the efficient treatment of the information and, consequently, it appears more appropiate to use the set of data relating to the decision-making unit, i.e. to obtain $(\partial p_i / \partial x_j)$ for each element of the sample and summarize the result in a single value by calculating the average of these derivatives. In short, we propose to emphasizing on the average behaviour of the sample, rather than the behaviour of «an average decision-making unit», in order to avoid a loss of information which does not appear to be justified. When the x_i vector includes dummy variables, then, according to our strategy, the quantification of expression (3) can be summarized by calculating p_i for each element and obtaining the average for each specific sub-sample distinguished by the value associated with the dummy variable (x_j); the difference between these average probabilities gives the above mentioned change in probability associated with a change in the variable x_j.

Our next objective is to carry out an analysis of both measurements, with the purpose of determining, if possible, the superiority of one over the other according to some kind of criterion. In order to establish this superiority, the criterion of «robustness» would appear to be appropriate. In our particular approach, we will consider that one measure is more robust than another if, in the face of small variations in the sample information, the value of the former varies less than the value of the latter. In order to examine this question, we have designed a simulation exercise which, using a logit and a probit model, generates different samples; specifically, 100 models are generated with a sample size of $N = 100$.

Denoting the derivative at the mean point and the mean of the derivatives as D_k and DM_k respectively, we set out to analyze the influence that the elimination of an element has on such values. With this aim, our robustness measure can be expressed as:

$$\frac{\sum_{i=1}^{100} (D_{ki} - VR_k)^2}{N} \quad \text{and} \quad \frac{\sum_{i=1}^{100} \left(DM_{ki} - VR'_k\right)^2}{N} \tag{4}$$

where k represents the explanatory variable for which the effect is calculated, VR and VR' denote, respectively, the value of D_k and DM_k for the whole sample, and D_{ki} and DM_{ki} are indicating the value of D_k and DM_k when the element i is eliminated.

The experiment concludes that, in all replications and in both models, D_k presents greater dispersion than DM_k. Furthermore, when calculating the mean of these dispersions, it can be observed in probit model that the dispersion of D_k is four times greater than that corresponding to DM_k. Therefore, the measure based on the average of the derivatives is more stable in the face of small changes in the sample information and, consequently, is more robust than the traditional method. This same general conclusion is obtained when we use an empirical application with real data.