Un estimador simple en dos etapas para un modelo de elección binaria con regresor entero endógeno

Mariana Marchionni[*]
Universidad Nacional de La Plata

Walter Sosa Escudero
Universidad de San Andrés

Resumen

Este trabajo discute un modelo de elección binaria con un regresor entero endógeno, y propone un método de estimación simple en dos etapas para obtener estimadores consistentes y asintóticamente normales de los parámetros de interés, en el espíritu de Rivers y Vuong (1988). La estrategia de estimación propuesta se evalúa a través de experimentos de Monte Carlo, de los que surge fuerte evidencia sobre su buen desempeño. Más aún, esta estrategia simple y natural genera estimaciones muy similares a las obtenidas por Weiss (1999) en su aplicación al modelo de aprobación de solicitudes de tarjetas de crédito.

Clasificación JEL: C13, C51, C35, C15
Palabras claves: variables enteras, probit endógeno, Monte Carlo, bootstrap

1 Introducción

Varias aplicaciones empíricas pueden representarse mediante modelos de elección binaria donde uno de los regresores es una variable entera endógena, provocando que las estrategias econométricas estándar, como los estimadores máximo verosímiles de los modelos probit o logit, se vuelvan inconsistentes. Un ejemplo discutido por Weiss (1999) consiste en un modelo de los determinantes de la aprobación de solicitudes para tarjetas de crédito, en donde una de las variables explicativas es el número de veces que el solicitante mantuvo un saldo deudor en otras cuentas de crédito. La endogeneidad de esta variable entera puede justificarse por la posible existencia de factores inobservables que afectan tanto el uso de fuentes de crédito por parte de los demandantes de tarjetas como la decisión de los encargados de evaluar las solicitudes. Un

[*] Departamento de Economía, Facultad de Ciencias Económicas, Universidad Nacional de La Plata, 6 entre 47 y 48, 5to Piso, Oficina 516. (1900) La Plata, Argentina. Tel/Fax: 54-221-422-9383. mariana@depeco.econo.unlp.edu.ar , wsosa@udesa.edu.ar. Una versión de este trabajo constituye el Capítulo III de la tesis doctoral de Mariana Marchionni (Marchionni, 2005), quien agradece la ayuda financiera brindada por la Academia Nacional de Ciencias Económicas. Los autores agradecen a Sebastián Galiani por sus valiosos comentarios y a William Greene por ceder gentilmente los datos sobre tarjetas de crédito utilizados en la aplicación empírica. Cualquier error que pueda subsistir es de exclusiva responsabilidad de los autores.

segundo ejemplo es un modelo de la decision de las mujeres de participar o no en el mercado laboral, que incluye al numero de hijos como regresor, y varios autores, como Schultz (1978), Dooley (1982), Moffitt (1984) o Hotz and Miller (1988), han argumentado eficazmente acerca de la posible endogeneidad de esta variable.

Weiss (1999) propone un modelo estructural donde la variable de interés es binaria y uno de los regresores es una variable entera endógenamente determinada. La estructura probabilística de este modelo se especifica completamente en forma paramétrica, de manera que todos los parámetros desconocidos pueden estimarse en forma consiste y asintoticamente eficiente por el método de máxima verosimilitud. Sin embargo, la compleja estructura de este modelo introduce una carga computacional considerable, que involucra procesos de integración numérica, lo que hace deseable explorar otras alternativas de estimación de implementación más sencilla.

El modelo propuesto por Weiss (1999) acomoda adecuadamente dos rasgos particulares de las variables involucradas: una variable dependiente binaria y un regresor endógeno entero. Rivers y Vuong (1988) consideran un caso más flexible en donde el regresor endógeno es continuo y su forma reducida puede representarse mediante una función lineal. En este contexto, puede implementarse un método simple en dos etapas basado en información limitada para obtener estimadores consistentes y asintóticamente normales de los parámetros de interés. A la luz de estos resultados, este trabajo explota el hecho de que la mayoría de los modelos para variables enteras son esencialmente modelos de regresión exponencial. Concentrádose en este aspecto del modelo, y no en la naturaleza entera de la variable endógena, se propone un estimador en dos etapas en el espíritu de la estrategia de Rivers y Vuong (1988), donde ahora la primera etapa del estimador involucra un problema de regresión no lineal. Asimismo, esta estrategia de estimación provee un test simple para evaluar la hipótesis nula de exogeneidad. La estrategia de estimación propuesta se evalúa a través de experimentos de Monte Carlo, de los que surge fuerte evidencia sobre su buen desempeño. Más aún, esta estrategia simple y natural produce estimaciones puntuales llamativamene similares a las obtenidas por Weiss (1999) en su aplicación al modelo de aprobación de solicitudes de tarjetas de crédito.

El resto del trabajo se organiza como sigue. En la sección 2 se presenta el modelo de interés y se discute el modelo propuesto por Weiss (1999). En la sección 3 se propone un estimador máximo verosímil condicional para el modelo de elección binaria considerando una especificación particular (la aditiva) para el modelo de la variable entera y se discuten otras especificaciones relacionadas que han sido estudiadas en la literatura. En la sección 4 se presenta evidencia de experimentos de Monte Carlo que permiten evaluar la performance en muestras chicas de ese estimador. La sección 5 discute una especificación multiplicativa para la variable endógena y propone otro estimador en dos etapas. En la sección 6 se presenta la aplicación empírica de aprobación de solicitudes para tarjetas de crédito estudiado por Weiss (1999) en base a un modelo de Greene (1994). Finalmente, en la sección 7 se concluye.

2 El modelo

El interés de este trabajo se centra en un modelo de elección binaria con un regresor entero endógeno. Una especificación general para la forma estructural de este modelo viene dada por las ecuaciones (1) y (2).

(1) $Y_{1i}^{*} = \delta\ Y_{2i} + X_{1i}\beta_1 + u_{1i}$ y $Y_{1i}=1[Y_{1i}^{*}>0]$

(2) $E[Y_{2i} \mid X_{2i}] = \exp(X_{2i}\beta_2)$,

,con $i = 1, \ldots, n$, Y_{1i}^* es una variable latente no observable de la que sólo se conoce su signo a través del indicador Y_{1i}, y Y_{2i} es una variable entera y potencialmente endógena con esperanza condicional exponencial, supuesto consistente con una amplia gama de modelos. X_{1i} y X_{2i} son, respectivamente, vectores de k_1 y k_2 regresores exógenos cuyo primer elemento es la unidad.

La especificación discutida por Weiss (1999) –en adelante *modelo de Weiss*– responde a la estructura general dada por las ecuaciones (1) - (2). En este modelo la variable entera viene determinada por las ecuaciones (3) y (4) debajo. La esperanza condicional de la variable entera –ecuación (3)– tiene la típica forma exponencial pero depende de un término inobservable e independiente de X_{2i}, representado por v_{2i}.[1] Este término indica la presencia de heterogeneidad no observada afectando los propios parámetros de la distribución condicional de la variable entera. La ecuación (4) implica que, condicional en la realización de v_{2i}, la variable entera sigue una distribución poisson con parámetro λ_i.[2]

(3) $$E[Y_{2i} \mid X_{2i}, v_{2i}] \equiv \lambda_i = \exp(X_{2i}\beta_2 + v_{2i})$$

(4) $$P(Y_{2i} = y_2 \mid v_{2i}) \equiv Po(\lambda_i, y_2) = \frac{\lambda_i^{y_2} \exp(-\lambda_i)}{y_2!} \quad , \ y_2 = 0, 1, 2, \ldots$$

Cuando v_{2i} tiene distribución exponencial gamma (EG), es posible encontrar una forma cerrada para la distribución marginal de la variable entera. Después de la normalización $E[\exp(v_{2i})]=1$, necesaria para la identificación de (3), la p.d.f de v_{2i} viene dada por

(5) $$v_{2i} \sim EG(\alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \exp(-\alpha \exp(v_{2i})) \exp(\alpha \ v_{2i})$$

donde α es el parámetro de la función gamma asociada a esta distribución y $\mathrm{Var}[\exp(v_{2i})]=1/\alpha$. La especificación dada por las ecuaciones (3), (4) y (5) corresponde al modelo binomial negativo para la variable entera, donde $E[Y_{2i}|X_{2i}] = \exp(X_{2i}\beta_2)$ y $\mathrm{Var}[Y_{2i}|X_{2i}]=\exp(X_{2i}\beta_2)[1+ \exp(X_{2i}\beta_2)/\alpha]$.

La representación que adopta el modelo de Weiss para incorporar la endogeneidad de Y_{2i} en la ecuación de elección binaria, viene dada por:

(6) $$u_{1i} = \varphi \ h(v_{2i}) + e_{1i} \qquad \text{con } h(v_{2i})=\Phi^{-1}[G(v_{2i})]$$

donde G(.) y Φ(.) son las funciones de distribución acumulada de v_{2i} y la normal estándar respectivamente, con lo que h(.) es normal estándar. Además, supone que la distribución de e_{1i} condicional en (Y_{2i}, v_{2i}, X_{1i}) es normal con media cero y varianza $\sigma_{e_1}^2 = 1 - \varphi^2$, y por lo tanto la distribución condicional de u_{1i} es normal estándar.

La obtención de estimadores máximo verosímiles de los parámetros del modelo de Weiss implica una dificultad computacional considerable, por lo que resulta conveniente evaluar la hipótesis nula de exogeneidad de la variable entera, es decir φ=0, antes de encarar la estimación conjunta del modelo. Si la endogeneidad no fuera relevante la ecuación de elección

---

[1] En Mullahy (1997) se discuten estimadores consistentes para el modelo de la esperanza condicional de una variable entera en presencia de correlación entre los regresores observables e inobservables.

[2] La incorporación de heterogeneidad no observable en la determinación de la media condicional de Y_{2i} a través del término v_{2i} permite romper con el supuesto de equidispersión –igualdad entre esperanza y varianza condicional de la variable entera– al que el modelo de poisson está atado.

binaria tendría la estructura de un modelo probit estándar, pudiéndose estimar, como es usual, por el método de máxima verosimilitud independientemente de la ecuación de la variable entera.[3] Weiss desarrolla un test para evaluar Ho:φ=0 versus la hipótesis alternativa $\varphi\neq0$ basado en Smith y Blundell (1986).[4]

El modelo de Weiss logra acomodar los dos rasgos fundamentales del modelo bajo consideración: la endogeneidad en un modelo de elección binaria y la naturaleza entera de la variable explicativa endógena. Sin embargo, la búsqueda de estimadores completamente eficientes en ese contexto implica un procedimiento computacionalmente muy complejo que requiere, entre otras cosas, de la aplicación de procesos de integración numérica, de manera que su replicación para encarar sistemáticamente el análisis empírico de problemas que respondan a la misma estructura es improbable o demasiado costosa.

En este trabajo se propone una estructura alternativa para el modelo de elección binaria con un regresor endógeno entero, basada en supuestos no más arbitrarios que los adoptados por Weiss pero de estimación considerablemente más simple. Nuevamente partimos de la especificación general del modelo dada por las ecuaciones (1) y (2). Como señala Mullahy (1997), la especificación exponencial para la media condicional implica que los modelos para variables enteras puedan considerarse esencialmente como modelos de regresión exponencial. En ese contexto, la única característica fundamental de la variable dependiente Y_{2i} sería la no negatividad y no el hecho de ser entera. Hay básicamente dos modelos de regresión exponencial consistentes con (2). El primero, al que llamaremos *modelo exponencial aditivo (MEA)*, viene dado por (7). El segundo, *modelo exponencial multiplicativo (MEM)*, viene dado por (8) y se caracteriza por un tratamiento simétrico de las variables observables (X_{2i}) y las inobservables (η_{2i}).

Modelo exponencial aditivo – MEA:

(7) $Y_{2i} = \exp(X_{2i}\beta_2) + u_{2i}$ con $E[u_{2i}|X_{2i}]=0$

---

[3] Otra situación en que la endogeneidad no sería relevante es cuando δ=0. Sin embargo, si $\varphi\neq0$ la distribución de u_{1i} podría no ser normal ya que depende de la especificación de $h(.)$ y de la distribución de e_{1i} –ver ecuación (6). En ese caso el estimador máximo verosímil del modelo probit resultaría inconsistente para el modelo de elección binaria, por lo que sería necesario estimar el modelo conjunto para determinar $h(v_{2i})$, la distribución de v_{2i} y si δ=0.

[4] Sustituyendo (6) en el modelo de elección binaria (1), y reemplazando $h(v_{2i})$ por el residuo generalizado que se define como

$$\hat{h}(v_{2i}) = E\left[h(v_{2i}) \mid Y_{2i} = y_2\right]\Big|_{(\hat{\beta}_2,\hat{\alpha})}$$
$$= \frac{1}{P(Y_{2i} = y_2)}\int h(v_{2i})Po(\hat{\lambda}_i, y_2)EG(v_{2i},\hat{\alpha})dv_{2i}$$

se obtiene

$$Y_{1i}^* = \delta\, Y_{2i} + X_{1i}\beta_1 + \varphi\, \hat{h}(v_{2i}) + e_{1i}$$

que responde a la especificación de un modelo probit estándar, pudiéndose estimar de la forma usual por máxima verosimilitud. El estadístico de prueba de este test es la razón t del coeficiente φ, cuya distribución asintótica bajo la hipótesis nula es normal estándar. Weiss también muestra cómo evaluar la hipótesis nula de exogeneidad de Y_{2i} usando un test de multiplicador de Lagrange (LM). Ambos tests son asintóticamente equivalentes bajo la hipótesis nula, pero el que se basa en Smith y Blundell es más fácil de implementar.

Modelo exponencial multiplicativo – MEM:

(8) $Y_{2i} = \exp(X_{2i}\beta_2 + \eta_{2i})$ con $E[\exp(\eta_{2i})|X_{2i}]=1$

Los modelos (7) y (8) son observacionalmente equivalentes a menos que se adopten supuestos adicionales sobre los momentos de orden superior de los términos de error.[5] En las secciones 0 y 4 se discute el modelo de elección binaria de la ecuación (1) cuando la especificación del regresor endógeno entero corresponde al MEA. Luego, en la sección 5, se estudia el mismo modelo cuando la variable endógena se determina a partir del MEM.

3 Estimación en dos etapas del modelo aditivo

En esta sección se propone una estrategia de estimación simple en dos etapas para los parámetros del modelo de elección binaria –ecuación (1)– cuando se adopta el modelo MEA para el regresor endógeno –ecuación (7). La endogeneidad es consecuencia de la correlación entre la variable entera, Y_{2i}, y el término de error de la ecuación de elección binaria (1), u_{1i}. Esta correlación se incorpora al modelo mediante la ecuación (9), que supone que u_{1i} y u_{2i} satisfacen:

(9) $u_{1i} = \theta\, u_{2i} + \varepsilon_{1i}$

Sustituyendo (9) en (1) se obtiene:

(10) $Y_{1i}^{*} = \delta\, Y_{2i} + X_{1i}\beta_1 + \theta\, u_{2i} + \varepsilon_{1i}$

Supongamos que se observa un muestra *iid* (Y_{1i}, Z_i) donde $Z_i \equiv (Y_{2i}, X_{1i}, u_{2i})$. Si la distribución de ε_{1i} condicional en Z_i es normal estándar, el modelo responde a la especificación de un probit estándar, de manera que los parámetros del modelo de elección binaria pueden ser estimados consistentemente por máxima verosimilitud de la manera usual.

En el caso relevante para nosotros, sin embargo, u_{2i} no es observable. Los estimadores máximo verosímiles del modelo probit de Y_{1i} en Y_{2i} y X_{1i} serán inconsistentes a menos que se adopten fuertes supuestos sobre la distribución de u_{2i} y sobre la posible dependencia entre u_{2i} y Y_{2i}. En este contexto se propone un estimador máximo verosímil condicional que se obtiene mediante un procedimiento en dos etapas que sigue en espíritu la estrategia de Rivers y Voung (1988) –en adelante RV. Llamaremos a este estimador 2EMEA por tratarse de un estimador en dos etapas para el modelo en que la variable endógena responde a la especificación MEA.

Manteniendo el supuesto en (7) que $E[u_{2i}|X_{2i}]=0$ y suponiendo que se observa una muestra *iid* (Y_{1i}, X_{1i}, Y_{2i}, X_{2i}), el siguiente procedimiento permite estimar consistentemente los parámetros δ, β_1 y θ:

1° etapa. Estimar los parametros de la ecuación (7) por el método de mínimos cuadrados no lineales (NLS). Obtener estimadores consistentes de u_{2i} definidos como $\hat{u}_{2i} \equiv Y_{2i} - \exp(X_{2i}\hat{\beta}_2)$. Aquí la propiedad de consistencia depende de la aditividad de u_{2i} y del supuesto $E[u_{2i}|X_{2i}]=0$.

[5] Notar que ambos modelos son consistentes con la esperanza condicional de Y_{2i} dada en (2) pero dan lugar a diferentes varianzas condicionales de Y_{2i}. Ver Wooldridge (1992).

2° etapa. Estimar por máxima verosimilitud un modelo probit de Y_{1i} en Y_{2i}, X_{1i} y $\hat{u}_{2i}$ como variables explicativas. Esta etapa consiste básicamente en reemplazar u_{2i} por el estimador consistente obtenido en la etapa previa. La propiedad de consistencia de los estimadores de δ, β_1 y θ depende de la normalidad de la distribución de ε_{1i} condicional en Z_i y de la correcta especificación de la media condicional de Y_{1i}^*.

La consistencia del estimador 2EMEA es trivial y depende estrictamente de la consistencia del estimador de la primera etapa. Conjeturalmente, y en base a los resultados de Rivers y Vuong (1988) la generalidad del estimador sugiere que el mismo deberia tener distribucion asintótica normal, lo que en este trabajo se confirmará a través de un extenso ejercicio de Monte Carlo. La derivación de resultados análiticos, si bien similares en espiritu al trabajo de Rivers y Vuong (1998) es una de las tareas que quedan pendientes en este trabajo, si bien los resultados experimentales son auspiciosos. Bajo normalidad asintotica es posible implementar tests *t* sobre los coeficientes individuales del modelo de elección binaria usando esa distribución asintótica. En particular, la hipótesis nula de exogeneidad de la variable entera puede evaluarse directamente mediante un test de significatividad individual sobre el coeficiente de los residuos de la primera etapa en la ecuación de la variable binaria. Si no puede rechazarse la hipótesis nula de exogeneidad, el modelo de elección binaria respondería a la especificación del modelo probit estándar y podría estimarse por máxima verosimilitud independientemente de la ecuación de la variable entera.

Con respecto a la matriz de varianzas y covarianzas del estimador 2EMEA, el estimador usual del probit que se obtiene en la segunda etapa resulta inconsistente porque ignora la variabilidad muestral de $\hat{\beta}_2$. En este trabajo se emplean técnicas de bootstrap para obtener estimaciones de los errores estándar de los coeficientes, dejando para investigaciones futuras la derivación analítica de la matriz de varianzas y covarianzas del estimador en dos etapas.

Otras especificaciones relacionadas discutidas en la literatura

Más arriba se señalaba que el estimador 2EMEA surge como una adaptación del método de máxima verosimilitud condicional en dos etapas propuesto por RV. El modelo estudiado por RV consiste en una ecuación estructural para la variable binaria de interés, que depende de un conjunto de regresores endógenos continuos representados por ecuaciones lineales en forma reducida. En la primera etapa del procedimiento se estiman por MCO los modelos de las variables endógenas y en la segunda etapa se estima un modelo probit estándar, incluyendo los errores de estimación de la etapa previa como regresores adicionales. Comparado con otros estimadores para modelos de elección binaria con regresores endógenos continuos, el estimador de RV se desempeña favorablemente en términos de performance en muestras chicas, utilidad para realizar tests de hipótesis y simplicidad en su implementación.

Como señalan RV, si bien el supuesto de normalidad conjunta en la distribución condicional de los errores es estándar para este tipo de modelos, no es necesario para la consistencia del estimador en dos etapas que proponen, así como tampoco lo es para el estimador 2EMEA. La consistencia sólo requiere que la distribución condicional de u_{1i} sea normal con media lineal en u_{2i} y varianza constante.

Otro estimador similar al de RV es el sugerido por Lee (1981) –probit con variables instrumentales (IVP). Este estimador también se obtiene mediante un procedimiento en dos etapas, pero se diferencia del de RV en que en la segunda etapa no se estima la forma estructural sino la forma reducida del modelo de elección binaria. Usando la notación para nuestro modelo, la segunda etapa del estimador IVP implicaría estimar un modelo probit estándar de Y_{1i} en X_{2i}, X_{1i} y el residuo estimado por mínimos cuadrados ordinarios en la primera etapa. Según la evidencia de Monte Carlo presentada por RV, el IVP se comporta algo peor en muestras chicas que el estimador de RV, mientras que no es posible establecer un ordenamiento en cuanto a su eficiencia asintótica.

Una especificación de nuestro modelo pero "cabeza abajo" es la que proponen Windmeijer y Santos Silva (1997). Puede pensarse que el modelo (1)-(2) surge como una de las dos triangularizaciones posibles del siguiente modelo simultáneo más general:

$$(11)\quad Y_{1i}^{*} = \delta\, Y_{2i} + X_{1i}\beta_1 + u_{1i} \quad \text{y} \quad Y_{1i}=1[Y_{1i}^{*} > 0]$$

$$(12)\quad E[Y_{2i} \mid X_{2i}] = \exp(\gamma\, Y_{1i} + X_{2i}\beta_2)$$

En el modelo (11)-(12) no puede darse que δ y γ sean simultáneamente distintos de cero. Esto violaría el requisito de consistencia Prob(Y_{1i}=1)+Prob(Y_{1i}=0)=1 (ver Maddala (1983, p.118) y Blundell y Smith (1994)). Por lo tanto la endogeneidad en este sistema no puede operar en ambas direcciones. El modelo relevante será entonces uno con variable dependiente entera y regresor endógeno binario (δ=0) o, alternativamente, otro con variable dependiente binaria y regresor endógeno entero (γ=0), como el discutido hasta ahora.

Windmeijer y Santos Silva (1997) se concentran en la otra triangularización, es decir, suponen que en la ecuación de la variable entera hay un regresor binario potencialmente endógeno, empleando alternativamente las especificaciones MEA y MEM para representar la variable entera. Estos autores discuten un estimador basado en el método generalizado de momentos (GMM). Intuitivamente, el procedimiento consiste en instrumentar la variable binaria endógena mediante la probabilidad predicha obtenida al estimar la ecuación (11), que responde a la estructura de un modelo probit estándar, por máxima verosimilitud. Sin embargo, pese a que los modelos MEA y MEM son observacionalmente indistinguibles, los instrumentos que resultan adecuados bajo una especificación generalmente dejan de serlo bajo la otra.[6]

4 Evidencia de Monte Carlo

En esta sección se presentan los resultados de algunos experimentos de Monte Carlo diseñados para evaluar las propiedades del estimador 2EMEA en muestras chicas. Si bien este es el principal objetivo, en cada caso se presentan también los resultados de usar el estimador máximo verosímil de un modelo probit estándar ignorando el problema de endogeneidad. La comparación de los resultados obtenidos por ambos métodos permite cuantificar la severidad de este problema cuando se utiliza ese estimador inconsistente.

Las simulaciones se basan en un modelo de elección binaria con dos regresores, uno de los cuales, Y_{2i}, es potencialmente endógeno y responde a la especificación MEA (ecuación (7)). El modelo viene dado por las ecuaciones (13) a (15).

$$(13)\quad Y_{1i}^{*} = \delta\, Y_{2i} + b_{11} + b_{12}x_{1i} + u_{1i} \quad \text{y} \quad Y_{1i}=1[Y_{1i}^{*} > 0]$$

$$(14)\quad Y_{2i} = \exp(b_{21} + b_{22}x_{2i}) + u_{2i}$$

$$(15)\quad u_{1i} = \theta\, u_{2i} + \varepsilon_{1i}$$

---

[6] La aplicación empírica que presentan Windmeijer y Santos Silva (1997) consiste en un modelo de determinación del número de visitas al médico en el último mes, que incluye entre los regresores un indicador binario sobre la percepción del propio estado de salud del individuo. En este caso puede pensarse que la endogeneidad surge como consecuencia de errores de medición de la variable binaria –autorreportada– posiblemente correlacionados con los determinantes inobservables de las visitas al médico.

Para las simulaciones se utilizaron los siguientes valores de los parámetros: δ=1, b_{11}=-2, b_{12}=1, b_{21}=-1 y b_{22}=1.[7] Cada simulación se basó en una muestra aleatoria de 1000 observaciones y se realizaron 500 replicaciones. La variable exógena x_{1i} se generó a partir de una distribución normal estándar, mientras que x_{2i} se obtuvo independientemente a partir de una distribución uniforme definida en el intervalo [0, $12^{1/2}$]. Los valores para $_{1i}$ y u_{2i} se generaron a partir de distribuciones normales independientes con medias cero y varianzas iguales a uno y σ_2^2, respectivamente. Las observaciones u_{1i} se generaron usando (15), con lo que E[u_{1i}]=0, Var[u_{1i}]$\equiv \sigma_1^2 = 1+\theta^2\sigma_2^2$ y $\theta = \rho\sigma_1/\sigma_2$, donde $\rho \equiv$ Corr[u_{1i}, u_{2i}]. Cuando $\rho \neq 0$ o, equivalentemente, $\theta \neq 0$, Y_{2i} es endógena en la ecuación (13) y los estimadores máximo verosímiles de un probit estándar de Y_{1i} en Y_{2i} y x_{1i} son inconsistentes para los parámetros –reescalados usando σ_1– del modelo de elección binaria. Este problema se agudiza a medida que ρ aumenta en valor absoluto, lo que, dada la varianza de u_{2i}, también implica mayor valor absoluto de θ.[8]

En primer lugar se realiza un experimento para estudiar el efecto de aumentos en la correlación entre los errores de la ecuación de elección binaria (u_{1i}) y de la variable entera (u_{2i}). La varianza de u_{1i} aumenta como consecuencia del aumento en la correlación porque σ_1^2=1/(1-ρ^2), mientras que la varianza de u_{2i} se mantiene fija (σ_2^2=1) a lo largo de los experimentos.Se usan cinco conjuntos alternativos de parámetros cuyos valores se presentan en las primeras tres columnas de la Tabla 1 –ver Apéndice. Estos valores de los parámetros permiten evaluar situaciones con correlaciones extremas como así también varios casos intermedios. Mientras la correlación aumenta desde cero a casi 0.9, el desvío estándar de u_1 más que se duplica, pasando de 1 a 2.24. Los resultados de Monte Carlo de estimar el modelo (13)-(15) con estos parámetros se presentan en las restantes columnas de la Tabla 1.

Naturalmente el sesgo del probit aumenta marcadamente con la correlación. Por ejemplo, la raíz cuadrada del error cuadrático medio (RMSE) del coeficiente de la variable endógena se triplica al pasar de una correlación nula a una moderada de 0.5, y casi se triplica nuevamente cuando la correlación alcanza el valor de 0.9. El sesgo para estos valores de correlación equivale a una sobreestimación del orden del 1.5%, 18% y 53%, respectivamente, del valor real del coeficiente (δ=1). Comportamientos similares se observan para los otros coeficientes.

Es de esperar que el sesgo del estimador 2EMEA también se acentúe a medida que la correlación aumenta. Intuitivamente, al reemplazar u_{2i} por su estimación existe parte del verdadero u_{2i} que continúa excluyéndose del modelo. Esa omisión produce un sesgo que se agudiza a medida que crece la correlación –y consecuentemente θ. A pesar de esto, es notable de esta tabla la buena performance del estimador 2EMEA, incluso para los niveles más altos de correlación. Por ejemplo, con una correlación de 0.9, el estimador 2EMEA sobrestima al verdadero coeficiente de la variable endógena en sólo un 2.2% y el RMSE del probit es 6 veces mayor que el RMSE del estimador 2EMEA, que sólo aumenta un 24% al pasar de una correlación nula a una de 0.9.

---

[7] El valor de b_{22} sólo se modifica para provocar cambios en la bondad del ajuste del modelo estimado en la primera etapa. Ver Tabla 2.

[8] El signo de θ no afecta los resultados de los experimentos de Monte Carlo aquí presentados pero sí los reportados por RV. Esta diferencia surge cuando se reescalan las estimaciones para poder comparar los distintos métodos. En el presente trabajo los estimadores del modelo probit estándar se reescalan usando SD(u_{1i})=$(1+\theta^2\sigma_2^2)^{1/2}$, mientras que RV usan $(1+(\theta+\delta)^2\ \sigma_2^2)^{1/2}$, que es el desvío estándar de la forma reducida del probit y que claramente no sólo depende de la magnitud sino también del signo de θ.

El buen desempeño del estimador 2EMEA que se desprende de la Tabla 1 en parte se debe a que la bondad del ajuste del modelo exponencial –R^2 implícito en el modelo no lineal y que por simplicidad será referido desde ahora como R^2– estimado en la primera etapa es muy alto. Intuitivamente, cuanto menor es la bondad del ajuste que surge de estimar la ecuación (15) por NLS, más se asemejan los residuos de esa regresión a la verdadera Y_{2i}, agravándose el problema de omitir u_{2i} y estimar el modelo como si se tratara de un modelo probit estándar. En relación al estimador 2EMEA, cuanto menor sea el R^2 del modelo de la primera etapa, más difícil será la identificación de δ y θ en la segunda, demorando la convergencia del estimador y aumentando las medidas de sesgo y RMSE para un tamaño de muestra dado.

Para explorar este punto y su efecto sobre el desempeño del estimador en muestras chicas, se realizan una serie de experimentos de Monte Carlo para distintos ajustes del modelo MEA, manteniendo constante la correlación entre u_{1i} y u_{2i}, y por ende también la propia varianza de u_{1i}. Dada la distribución de x_{2i} y los parámetros b_{21} y b_{22}, el R^2 del modelo de la primera etapa aumenta cuando cae la varianza del término de error de ese modelo, es decir cuando cae σ_2^2. Por ejemplo, para los experimentos presentados en la Tabla 1, la varianza de u_{2i} es relativamente baja (σ_2^2=1) y el ajuste promedio muy alto (95%). Por otra parte, también θ aumenta cuando cae σ_2^2, por lo que la varianza de u_{2i} no sólo afecta el ajuste del modelo para la variable endógena sino que también modifica la relación entre u_{1i} y u_{2i}.

Una forma de provocar cambios en el R^2 sin modificar θ es mediante cambios en la varianza del término exponencial de la ecuación (15). Esta es la estrategia que se sigue en este ejercicio. Se realizan siete experimentos con una correlación entre moderada y alta de ρ=0.7, σ_2^2=1 y distintos valores de b_{22} que permiten que el R^2 promedio varíe entre un 12% y un 99%. Los valores de los parámetros empleados y los resultados de Monte Carlo de estimar el modelo (13)-(15) con estos parámetros se presentan en la Tabla 2 del Apéndice.

En la Tabla 2 puede verse cómo mejora la performance del estimador 2EMEA a medida que lo hace el ajuste del modelo estimado en la primera etapa. Cuando el ajuste es muy bajo, por ejemplo 12% en las últimas dos filas de la tabla, el desempeño de los estimadores 2EMEA, especialmente para estimar el coeficiente vinculado a la variable endógena –δ–, es muy pobre. Esto se manifiesta en la magnitud del sesgo, pero aún más en los RMSE, considerablemente superiores a los del probit a causa del explosivo aumento de la varianza del estimador en 2 etapas.

Para los experimentos previos los errores del modelo de la variable endógena, u_{2i}, se supusieron normales. Como se remarcó en la sección 0, la normalidad de u_{2i} no es condición necesaria para la consistencia del estimador 2EMEA. A continuación se realiza una serie de experimentos de Monte Carlo con el objetivo de evaluar la performance de este estimador ante cambios en la correlación entre los errores cuando los correspondientes a la ecuación de la variable endógena se distribuyen uniformemente en el intervalo (-1.73, 1.73), de manera que como en las Tablas 1 y 2, σ_2^2=1. En la Tabla 3 del Apéndice se explora el efecto de aumentos en la correlación de 0 a 0.9, manteniendo constante el R^2 promedio del modelo estimado en la primera etapa. En la Tabla 4 se explora el efecto de cambios en el ajuste de ese modelo para una correlación entre moderada y alta de ρ=0.7. Como era de esperar, los resultados son similares a los obtenidos para errores normales en la ecuación de la variable endógena –ver Tablas 1 y 2.

Por último, se presenta un experimento que tiene por objetivo evaluar, en muestras chicas, la potencia del test de exogeneidad de Y_{2i} en el modelo de elección binaria, para distintos valores del parámetro θ, y por ende de la correlación entre los términos de error. La hipótesis nula de exogeneidad (H_0: θ=0) se evalúa mediante el estadístico de prueba z, definido como $z \equiv \hat{\theta} / \hat{sd}(\hat{\theta})$, cuya distribución asintótica bajo la hipótesis nula es N(0,1). Se computa la frecuencia relativa de rechazo de H_0 utilizando para el contraste esa distribución considerando distintos niveles de significatividad asintótica y bondades de ajuste del modelo estimado en la primera etapa. En cada caso se estima el error estándar del coeficiente estimado $\hat{\theta}$ mediante técnicas de bootstrap, usando 100 submuestras para cada una de las 500 replicaciones del experimento de Monte Carlo.

En la Tabla 5 se presentan los resultados de este ejercicio. En las primeras tres columnas se reportan los valores de los parámetros empleados para cada experimento. En las demás columnas se muestran los valores promedio en 500 replicaciones del coeficiente estimado $\hat{\theta}$ y de su desvío estándar estimado, las frecuencias relativas de rechazo de la hipótesis nula de exogeneidad y el p-valor asociado al estadístico de Kolmogorov-Smirnov (K-S) para evaluar la hipótesis de normalidad de la distribución empírica de $\hat{\theta}$. La hipótesis nula de exogeneidad se evalúa para tres niveles de significatividad asintótica alternativos y considerando siempre una hipótesis alternativa bilateral. La primera fila de la tabla corresponde al caso en que la hipótesis nula de exogeneidad es cierta, por lo que las frecuencias relativas de rechazo que se reportan constituyen el error tipo I del test. Si en lugar de tratarse de la distribución asintótica del estadístico z, la normal estándar fuera su distribución exacta, el error tipo I coincidiría con el correspondiente nivel de significatividad del test. Para los demás casos la frecuencia relativa de rechazo de H_0 mide la potencia del test.

Se desprende de la Tabla 5 que aún para correlaciones entre bajas y medias –alrededor de ρ=0.4– la potencia del test es muy alta. Por su parte, la probabilidad de rechazar una hipótesis nula de exogeneidad cuando es cierta se acerca mucho al nivel de significatividad asintótico del test, en especial para el caso con menor R^2.

Según los resultados de aplicar el test de normalidad de Kolmogorov-Smirnov, no puede rechazarse la normalidad de la distribución empírica del estimador 2EMEA de θ para los niveles de significatividad habituales –la única excepción aparece en la última fila de la tabla.

5 Un modelo multiplicativo para la variable endógena

Hasta aquí el trabajo se ha concentrado en la especificación aditiva –MEA– del modelo de regresión para la variable endógena, representado por la ecuación (7). Como se señaló anteriormente, haciendo los supuestos necesarios sobre la esperanza del término de error, tanto el MEA como el MEM –Modelo Exponencial Multiplicativo– son consistentes con $E[Y_{2i} \mid X_{2i}] = \exp(X_{2i}\beta_2)$ y observacionalmente indistinguibles entre sí.

En esta sección se discute brevemente el modelo de elección binaria cuando el regresor entero endógeno viene determinado por el MEM – ecuación (8)– y se propone otro estimador máximo verosímil condicional en dos etapas que resulta consistente bajo esta nueva especificación, y que al igual que el estimador 2EMEA es de muy sencilla aplicación. El modelo viene dado por las ecuaciones (16)-(18).

(16) $Y_{1i}^* = \delta\, Y_{2i} + X_{1i}\beta_1 + u_{1i}$ y $Y_{1i}=1[Y_{1i}^*>0]$

(17) $Y_{2i} = \exp(X_{2i}\beta_2 + \eta_{2i})$ con $E[\exp(\eta_{2i})|X_{2i}]=1$

(18) $\mu_{1i} = \theta\, (\eta_{2i} - M) + \varepsilon_{1i}$ con $E[\eta_{2i}|X_{2i}]=M$

La diferencia con el MEA es que el término de error en la ecuación (17) recibe un tratamiento simétrico al de las variables observables X_{2i}. La ecuación (18) introduce la endogeneidad de Y_{2i} en el modelo de elección binaria.[9] Se supone que la distribución condicional de ε_{1i} es normal estándar.

Nuevamente, si θ=0 los parámetros –reescalados– de la ecuación (16) pueden estimarse consistentemente mediante el estimador máximo verosímil de un probit estándar de Y_{1i} en Y_{2i} y X_{2i}. Cuando $\theta \neq 0$ tanto este método como el estimador 2EMEA resultan inconsistentes. El siguiente procedimiento en dos etapas permite estimar consistentemente los parámetros δ, β_1 y θ cuando el regresor endógeno entero en el modelo de elección binaria sigue la especificación MEM.

1° etapa. Linealizar la ecuación (17) aplicando logaritmos y estimar sus parametros por el método de mínimos cuadrados ordinarios (MCO). Obtener los residuos de esa regresión. Estos residuos son estimadores consistentes de η_{2i}-M.

2° etapa. Estimar por máxima verosimilitud un modelo probit estándar de Y_{1i} en Y_{2i}, X_{2i} y los residuos obtenidos en la etapa anterior.

A continuación se presentan los resultados de un experimento de Monte Carlo diseñado para evaluar la performance en muestras chicas del estimador 2EMEM –2 Etapas Modelo Exponencial Multiplicativo. Las simulaciones para el experimento de Monte Carlo se basan en el modelo (19)-(21).

(19) $Y_{1i}^{*} = \delta\ Y_{2i} + b_{11} + b_{12}x_{1i} + u_{1i}$ y $Y_{1i}=1[Y_{1i}^{*}>0]$

(20) $Y_{2i} = \exp(b_{21} + b_{22}x_{2i} + \eta_{2i})$ con $E[\exp(\eta_{2i})|x_{2i}]=1$

(21) $\mu_{1i} = \theta\ (\eta_{2i} - M) + \varepsilon_{1i}$ con $E[\eta_{2i}|X_{2i}]=M$

En general, se emplearon los mismos valores de los parámetros que en los ejercicios de la sección 4, es decir δ=1, b_{11}=-2, b_{12}=1, b_{21}=-1 y b_{22}=1. Al igual que antes cada simulación se basó en una muestra aleatoria de 1000 observaciones y se realizaron 500 replicaciones. x_{1i} y $_{1i}$ se generaron, independientemente, a partir distribuciones normal estándar, mientras que para x_{2i} se usó una distribución uniforme definida en el intervalo [0, $12^{1/2}$]. Los valores para η_{2i} se generaron a partir de una distribución normal con media M<0 y varianza σ_η^2=-2M, de manera que $\exp(\eta_{2i})$ sigue una distribución log-normal que satisface $E[\exp(\eta_{2i})|X_{2i}]=1$. De la ecuación (21), se sigue que la distribución de u_{1i} tiene media igual a cero y varianza igual a $\sigma_1^2 = 1 + \theta^2\sigma_2^2$, y que $\theta=\rho\sigma_1/\sigma_2$, donde $\rho \equiv \mathrm{Corr}[u_{1i}, \eta_{2i}]$.

En primer lugar se fijó M=-0.02, que implica un R^2 –promedio en 500 replicaciones– del 96% para el modelo lineal estimado en la primera etapa por MCO, y se repitió el experimento para distintos niveles de correlación entre los residuos. Se estimó el modelo (19)-(21) usando el estimador máximo verosímil del modelo probit estándar, el estimador 2EMEA y el método 2EMEM. Los resultados de este ejercicio se presentan en la Parte A de la Tabla 6, mientras que en la Parte B se muestran los correspondientes a M=-1 con un ajuste promedio del modelo de la primera etapa del 33%. Los resultados sugieren una buena performance del estimador 2EMEM tanto en términos absolutos como relativos, aún cuando hay considerable endogeneidad en el modelo o el ajuste de la primera etapa es sólo moderado.

[9] Notar que el término η_{2i}-M cumple aquí un rol similar al de la función h(.) en la ecuación (6) del modelo de Weiss (1999). La necesidad de encontrar una expresión cerrada para la distribución de Y_{2i} es lo que hace que Weiss utilice $h(v_{2i})$ con v_{2i} exponencial gamma y no η_{2i}-M con η_{2i} normal. Ver Weiss (1999), pág. 432-433.

6 Una aplicación empírica: modelo de aprobación de solicitudes para la obtención de tarjetas de crédito.

Como aplicación empírica se presenta aquí un modelo de aprobación de solicitudes para obtener tarjetas de crédito. Los datos fueron empleados originalmente por Greene (1994) y luego por Weiss (1999), quién estimó el modelo por Máxima Verosimilitud empleando la especificación presentada en la sección 2.

La variable dependiente del modelo de elección binaria indica si la solicitud para una nueva tarjeta de crédito fue o no aprobada. Se supone que esta decisión es afectada por una variable entera (*Major Derogatory Reports* - *MDR*), que representa el número de veces que el solicitante mantuvo una deuda en alguna cuenta de crédito por más de sesenta días. La muestra cuenta con información sobre 1319 solicitudes, de las cuales 1023 fueron aprobadas.

Las empresas de tarjetas de crédito tercerizan la evaluación de las solicitudes. Los equipos que finalmente se encargan de esta instancia han automatizado el proceso que permite obtener la calificación crediticia del solicitante –que a su vez determinará si la solicitud es o no aprobada–, pero ese es un proceso que se mantiene en secreto. Greene (1994) encuentra que el número de *MDR* es una variable altamente significativa para explicar la aprobación de solicitudes para obtener una nueva tarjeta de crédito, a la vez que sugiere que podría ser apropiado un modelo simultáneo. La existencia de factores inobservables que afecten tanto al uso del crédito por parte de los individuos como a las decisiones sobre la aprobación de sus solicitudes daría lugar a la endogeneidad de *MDR* en la ecuación de elección binaria. Los modelos de elección binaria y para la variable entera –*MDR*– en Weiss son básicamente iguales a los que presenta Greene (1994), excepto porque Weiss incluye *MDR* y MDR^2 como regresores del primero. La definición de las variables independientes se da en la Tabla 7 del Apéndice, donde se mantienen las denominaciones originales de las mismas.

Además del modelo binomial negativo para la variable endógena, Weiss explora otras especificaciones alternativas, en particular un modelo binomial negativo con exceso de ceros –*zero-inflated negative binomial*. En la aplicación de aprobación de solicitudes evalúa comparativamente ambos modelos y concluye que no hay evidencia contundente que permita elegir uno por sobre el otro.

En la Tabla 8 se presentan los resultados de estimar el modelo de aceptación de solicitudes para tarjetas de crédito usando alternativamente el estimador del modelo probit estándar, el estimador del modelo de Weiss –se reportan los resultados que aparecen en Weiss (1999)–, el estimador 2EMEA y el estimador 2EMEM. El resultado más interesante es la llamativa similitud de los coeficientes estimados de *MDR* y MDR^2 usando los estimadores máximo verosímiles del modelo de Weiss y el estimador 2EMEM. En ambos casos también se rechaza la hipótesis nula de exogeneidad de *MDR* en el modelo de elección binaria. En cuanto a las demás variables, los intervalos de confianza de los coeficientes estimados por máxima verosimilitud a partir del modelo de Weiss siempre contienen a los estimados por 2EMEM, y viceversa.

Por su parte los resultados de utilizar el estimador 2EMEA difieren sustancialmente de los demás. Si bien también se rechaza la exogeneidad de *MDR*, las estimaciones del coeficiente de esta variable, de su cuadrado y del coeficiente θ no están contenidas en los intervalos de confianza para las estimaciones de Weiss ni para las que surgen usando 2EMEM. Lo mismo puede decirse para muchas de las variables explicativas independientes. Determinar cuál de las tres especificaciones del modelo endógeno mejor se adapta a los datos escapa a los objetivos de este trabajo, por lo que se deja para futuras investigaciones.

7 Conclusiones

En este trabajo se discutió la estimación de un modelo de elección binaria donde una de las variables explicativas es endógena y entera. El único trabajo –al menos del único que tenemos conocimiento– que se ha concentrado en la estimación de un modelo con esta misma estructura es Weiss (1999), que propone una especificación particular para este modelo en un contexto de máxima verosimilitud. La búsqueda de estimadores completamente eficientes conlleva la necesidad de adoptar un conjunto de supuestos ciertamente arbitrarios y una considerable dificultad computacional.

En este trabajo se exploró una alternativa de estimación consistente y de más sencilla aplicación para un modelo que responde a la misma estructura general que el modelo de Weiss, es decir, un modelo de elección binaria con regresor entero endógeno. La especificación particular que se adoptó –no más arbitraria que la propuesta por Weiss– surge de considerar que los modelos para variables enteras son en esencia modelos de regresión exponencial. Se propusieron así dos estimadores máximo verosímiles condicionales que se obtienen en dos etapas: el estimador 2EMEA cuando la especificación del término de error en el modelo de la variable entera es aditiva y el 2EMEM cuando es multiplicativa. Estos estimadores son consistentes aún cuando los errores en la ecuación de la variable entera no son normales y su cómputo es muy simple a partir de cualquiera de los paquetes econométricos disponibles.

Los experimentos de Monte Carlo, por su parte, mostraron un buen desempeño –en términos de sesgo y RMSE– de estos estimadores en muestras chicas, aún con grados de endogeneidad considerables y ante distribuciones de los errores en la ecuación de la variable entera significativamente alejadas de la normal. También los experimentos de Monte Carlo mostraron en qué medida la bondad del ajuste del modelo de la variable endógena es un importante determinante del desempeño del estimador en dos etapas.

Por último, se utilizaron los estimadores propuestos para estimar el modelo de aprobación de solicitudes para tarjetas de crédito discutido por Weiss (1999). Al comparar resultados, sorprende la similitud de los obtenidos por Weiss a partir del estimador máximo verosímil de su modelo con los que surgen de emplear el estimador en dos etapas 2EMEM.

Varios aspectos quedan pendientes para investigaciones futuras. En primer lugar, la formalización de algunos de los resultados obtenidos experimentalmente y la derivación analítica de las matrices de varianzas y covarianzas de los estimadores en dos etapas –estimadas aquí empleando técnicas de bootstrap- es un paso natural, a la luz de los auspiciosos resultados experimentales de este trabajo. En segundo lugar, resulta de interés investigar otras aplicaciones, como el caso de fertilidad en las ecuaciones de participación laboral discutido en la Introduccion. Los pasos en dicha direccion son todavia incipientes, ver Marchionni (2005), si bien relevantes.

Referencias

Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

Dooley, M. D. (1982). Labor Supply and Fertility of Married Women: An Analysis with Grouped and Individual Data from the U.S. Census. *Journal of Human Resources*, 17, pp. 499-532.

Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. *Working Paper EC*-94-10, Department of Economics, New York University.

Hotz, V.J. and R. A. Miller. (1988). An Empirical Amalysis of Life Cycle Fertility and Female Labor Supply. *Econometrica*, 56, pp. 91-118.

Lee, L. F. (1981). Simultaneous equation models with discrete and censored dependent variables. En: C. Manski y McFadden, eds. *Structural analysis of discrete data with economic applications.* MIT Press, Cambridge.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Econometric Society Monographs. Cambridge University Press, Cambridge.

Marchionni, M. (2005). Estructura Demografica e Ingresos. Un Analisis Empirico para Argentina, tesis doctoral no publicada, Universidad Nacional de La Plata.

Moffitt, R. (1984). Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model. *Review of Economic Studies*, 51, pp. 263-278.

Mullahy, J. (1997). Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior. *The Review of Economics y Statistics.* Vol. 79, No. 4.

Rivers, D. and Q. H. Vuong. (1988). Limited Information Estimators and Exogeneity Tests for Simultaneous probit Models. *Journal of Econometrics.* Vol 39, pp. 347-366.

Schultz, T. P. (1978). The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimates". En R. Ehrenberg (ed.). *Research in Labor Economics*. Vol. 2, JAL Press, pp. 273-351.

Smith, R. and Richard W. Blundell (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply. *Econometrica*, Vol. 54, No.3, pp. 679-685.

Weiss, A. (1999). A Simultaneous Binary Choice / Count Model with an Application to Credit Card Approvals. In *Cointegration, Causality, and Forecasting: A Festchrift in honour of Clive W. J. Granger.* Oxford University Press.

Windmeijer, F. A. G. and J. M. C. Santos Silva. (1997). "Endogeneity in Count Data Models: An Application to Demand for Health Care". *Journal of Applied Econometrics.* Vol. 12, No. 3, pp. 281-294.

Wooldridge, J. M. (1992). Some Alternatives to the Box-Cox Regression Model. *International Economic Review.* Vol. 33, No. 4, pp. 935-955.

Wooldridge, J. M. (1997). Quasi-Likelihood Methods for Count Data. *Handbook of Applied Econometrics,* Vol. 2, Chapter 8.

Apéndice

Tabla 1

Efecto de aumentos en la correlación

Parámetros			Estimador	δ			b_{11}			b_{12}			R^2
θ	ρ	σ_2		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
0.0	0.0	1.0	Probit	1.015	0.015	0.003	-2.031	-0.031	0.007	1.020	0.020	0.004	-
			2EMEA	1.017	0.017	0.003	-2.035	-0.035	0.007	1.022	0.022	0.004	0.95
0.1	0.1	1.0	Probit	1.044	0.044	0.004	-2.074	-0.074	0.007	1.019	0.019	0.004	-
			2EMEA	1.017	0.017	0.003	-2.033	-0.033	0.007	1.020	0.020	0.004	0.95
0.5	0.4	1.0	Probit	1.179	0.179	0.009	-2.305	-0.305	0.015	1.042	0.042	0.005	-
			2EMEA	1.017	0.017	0.003	-2.034	-0.034	0.007	1.021	0.021	0.004	0.95
1.0	0.7	1.0	Probit	1.348	0.348	0.016	-2.629	-0.629	0.029	1.081	0.081	0.006	-
			2EMEA	1.022	0.022	0.004	-2.046	-0.046	0.008	1.024	0.024	0.005	0.95
2.0	0.9	1.0	Probit	1.530	0.530	0.024	-3.086	-1.086	0.050	1.095	0.095	0.007	-
			2EMEA	1.022	0.022	0.004	-2.052	-0.052	0.010	1.026	0.026	0.005	0.95

Notar: los valores presentados en las filas rotuladas *probit* corresponden a las estimaciones obtenidas usando el estimador máximo verosímil del modelo probit estándar. A fines comparativos se reescalan esos coeficientes multiplicándolos por el desvío estándar de $u_1 = (1-\rho^2)^{1/2}$.

Tabla 2

Efecto del ajuste del modelo estimado en la primera etapa

Parámetros				Estimador	δ			b_{11}			b_{12}			R^2
θ	ρ	σ_2	b_{22}		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
1.0	0.7	1.0	1.5	Probit	1.345	0.345	0.016	-2.631	-0.631	0.030	1.085	0.085	0.007	-
				2EMEA	1.027	0.027	0.004	-2.059	-0.059	0.009	1.031	0.031	0.005	0.998
1.0	0.7	1.0	1.0	Probit	1.348	0.348	0.016	-2.629	-0.629	0.029	1.081	0.081	0.006	-
				2EMEA	1.022	0.022	0.004	-2.046	-0.046	0.008	1.024	0.024	0.005	0.952
1.0	0.7	1.0	0.8	Probit	1.434	0.434	0.020	-2.755	-0.755	0.035	1.097	0.097	0.006	-
				2EMEA	1.018	0.018	0.003	-2.037	-0.037	0.007	1.018	0.018	0.004	0.861
1.0	0.7	1.0	0.6	Probit	1.997	0.997	0.045	-3.381	-1.381	0.062	1.227	0.227	0.011	-
				2EMEA	1.013	0.013	0.005	-2.036	-0.036	0.008	1.022	0.022	0.004	0.671
1.0	0.7	1.0	0.4	Probit	2.614	1.614	0.073	-3.723	-1.723	0.078	1.373	0.373	0.017	-
				2EMEA	1.004	0.004	0.011	-2.022	-0.022	0.010	1.020	0.020	0.004	0.422
1.0	0.7	1.0	0.2	Probit	2.832	1.832	0.082	-3.586	-1.586	0.072	1.427	0.427	0.020	-
				2EMEA	0.898	-0.102	0.038	-1.965	0.035	0.021	1.017	0.017	0.004	0.227
1.0	0.7	1.0	0.01	Probit	2.863	1.863	0.084	-3.391	-1.391	0.063	1.432	0.432	0.020	-
				2EMEA	2.353	1.353	2.428	-2.467	-0.467	0.926	1.015	0.015	0.004	0.124

Notar: los valores presentados en las filas rotuladas *probit* corresponden a las estimaciones obtenidas usando el estimador máximo verosímil del modelo probit estándar. A fines comparativos se reescalan esos coeficientes multiplicándolos por el desvío estándar de $u_1 = (1-\rho^2)^{1/2}$.

Tabla 3
Errores uniformes en la ecuación de la variable endógena

Parámetros			Estimador	δ			b_{11}			b_{12}			R^2
θ	ρ	σ_2		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
0.0	0.0	1.0	Probit	1.012	0.012	0.003	-2.025	-0.025	0.007	1.017	0.017	0.004	-
			2EMEA	1.014	0.014	0.003	-2.028	-0.028	0.007	1.019	0.019	0.004	0.95
0.1	0.1	1.0	Probit	1.045	0.045	0.004	-2.077	-0.077	0.008	1.017	0.017	0.004	-
			2EMEA	1.015	0.015	0.003	-2.031	-0.031	0.007	1.019	0.019	0.004	0.95
0.5	0.4	1.0	Probit	1.193	0.193	0.009	-2.320	-0.320	0.016	1.031	0.031	0.005	-
			2EMEA	1.019	0.019	0.004	-2.040	-0.040	0.008	1.022	0.022	0.005	0.95
1.0	0.7	1.0	Probit	1.360	0.360	0.017	-2.624	-0.624	0.029	1.030	0.030	0.005	-
			2EMEA	1.023	0.023	0.004	-2.050	-0.050	0.008	1.023	0.023	0.005	0.95
2.0	0.9	1.0	Probit	1.527	0.527	0.024	-3.039	-1.039	0.047	0.967	-0.033	0.006	-
			2EMEA	1.021	0.021	0.004	-2.051	-0.051	0.010	1.020	0.020	0.005	0.95

Notar: los valores presentados en las filas rotuladas *probit* corresponden a las estimaciones obtenidas usando el estimador máximo verosímil del modelo probit estándar. A fines comparativos se reescalan esos coeficientes multiplicándolos por el desvío estándar de $u_1 = (1\text{-}\rho^2)^{1/2}$.

Tabla 4
Ajuste del modelo de la variable endógena con errores uniformes

Parámetros				Estimador	δ			b_{11}			b_{12}			R^2
θ	ρ	σ_2	b_{22}		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
1.0	0.7	1.0	1.5	Probit	1.357	0.357	0.017	-2.621	-0.621	0.030	1.037	0.037	0.006	-
				2EMEA	1.028	0.028	0.005	-2.063	-0.063	0.010	1.035	0.035	0.006	0.99
1.0	0.7	1.0	1.0	Probit	1.360	0.360	0.017	-2.624	-0.624	0.029	1.030	0.030	0.005	-
				2EMEA	1.023	0.023	0.004	-2.050	-0.050	0.008	1.023	0.023	0.005	0.95
1.0	0.7	1.0	0.8	Probit	1.429	0.429	0.020	-2.722	-0.722	0.033	1.034	0.034	0.004	-
				2EMEA	1.018	0.018	0.004	-2.041	-0.041	0.008	1.019	0.019	0.004	0.86
1.0	0.7	1.0	0.6	Probit	2.112	1.112	0.050	-3.502	-1.502	0.068	1.195	0.195	0.010	-
				2EMEA	1.014	0.014	0.005	-2.038	-0.038	0.008	1.021	0.021	0.004	0.67
1.0	0.7	1.0	0.4	Probit	2.666	1.666	0.075	-3.755	-1.755	0.079	1.374	0.374	0.018	-
				2EMEA	1.014	0.014	0.011	-2.029	-0.029	0.011	1.019	0.019	0.004	0.42
1.0	0.7	1.0	0.2	Probit	2.841	1.841	0.083	-3.596	-1.596	0.072	1.427	0.427	0.020	-
				2EMEA	0.920	-0.080	0.040	-1.981	0.019	0.022	1.017	0.017	0.004	0.23
1.0	0.7	1.0	0.0	Probit	2.870	1.870	0.084	-3.403	-1.403	0.064	1.433	0.433	0.020	-
				2EMEA	-1.909	-2.909	3.429	-0.916	1.084	1.324	1.016	0.016	0.004	0.12

Notar: los valores presentados en las filas rotuladas *probit* corresponden a las estimaciones obtenidas usando el estimador máximo verosímil del modelo probit estándar. A fines comparativos se reescalan esos coeficientes multiplicándolos por el desvío estándar de $u_1 = (1\text{-}\rho^2)^{1/2}$.

Tabla 5

Potencia del test de exogeneidad

Parámetros			R^2 = 0.67 (promedio en 500 replicaciones) Estimaciones (promedio en 500 reps.)		Frec. rel. de rechazo de Ho: θ=0 Significatividad asintótica			K-S** p-valor	R^2 = 0.95 (promedio en 500 replicaciones) Estimaciones (promedio en 500 reps.)		Frec. rel. de rechazo de Ho: θ=0 Significatividad asintótica			K-S** p-valor
θ	ρ	σ_2	θ	Bootstraped SD(θ)	1%	5%	10%		θ	Bootstraped SD(θ)	1%	5%	10%	
0.0	0.0	1.0	0.0067	0.089	0.012	0.046	0.096	0.837	0.0018	0.077	0.004	0.036	0.098	0.531
0.1	0.1	1.0	0.1086	0.091	0.078	0.204	0.336	0.987	0.1022	0.078	0.090	0.274	0.394	0.758
0.5	0.4	1.0	0.5084	0.103	0.992	0.998	1.000	0.988	0.5081	0.086	1.000	1.000	1.000	0.737
1.0	0.7	1.0	1.0168	0.128	1.000	1.000	1.000	0.623	1.0119	0.106	1.000	1.000	1.000	0.656
2.0	0.9	1.0	2.0362	0.204	1.000	1.000	1.000	0.801	2.0492	0.178	1.000	1.000	1.000	0.029

*Test de normalidad Kolmogorov-Smirnov. Los p-valores corresponden a la hipótesis nula que postula que la distribución empírica del estimador de θ es normal.

Tabla 6

Parte A: Modelo Multiplicativo –R^2alto

Parámetros				Estimador	δ			b_{11}			b_{12}			R^2
θ	*ρ*	*M*	σ_2		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
0.0	0.0	-0.02	0.20	Probit	1.013	0.013	0.003	-2.025	-0.025	0.006	1.016	0.016	0.004	-
				2EMEA	1.015	0.015	0.003	-2.030	-0.030	0.006	1.018	0.018	0.004	0.96
				2EMEM	1.014	0.014	0.003	-2.029	-0.029	0.006	1.018	0.018	0.004	0.96
0.5	0.1	-0.02	0.20	Probit	1.019	0.019	0.003	-2.041	-0.041	0.006	1.016	0.016	0.004	-
				2EMEA	1.029	0.029	0.003	-2.041	-0.041	0.006	1.017	0.017	0.004	0.96
				2EMEM	1.014	0.014	0.003	-2.028	-0.028	0.006	1.018	0.018	0.004	0.96
1.0	0.2	-0.02	0.20	Probit	1.025	0.025	0.003	-2.057	-0.057	0.007	1.015	0.015	0.004	-
				2EMEA	1.043	0.043	0.004	-2.048	-0.048	0.007	1.012	0.012	0.004	0.96
				2EMEM	1.014	0.014	0.003	-2.027	-0.027	0.006	1.018	0.018	0.004	0.96
2.0	0.4	-0.02	0.20	Probit	1.038	0.038	0.004	-2.090	-0.090	0.007	1.014	0.014	0.004	-
				2EMEA	1.067	0.067	0.005	-2.049	-0.049	0.007	0.992	-0.008	0.004	0.96
				2EMEM	1.015	0.015	0.003	-2.029	-0.029	0.007	1.019	0.019	0.004	0.96
3.0	0.5	-0.02	0.20	Probit	1.050	0.050	0.004	-2.127	-0.127	0.008	1.014	0.014	0.004	-
				2EMEA	1.085	0.085	0.005	-2.033	-0.033	0.007	0.961	-0.039	0.004	0.96
				2EMEM	1.016	0.016	0.003	-2.033	-0.033	0.007	1.020	0.020	0.004	0.96
5.0	0.7	-0.02	0.20	Probit	1.073	0.073	0.005	-2.196	-0.196	0.011	1.012	0.012	0.004	-
				2EMEA	1.096	0.096	0.006	-1.948	0.052	0.007	0.875	-0.125	0.007	0.96
				2EMEM	1.020	0.020	0.004	-2.040	-0.040	0.008	1.021	0.021	0.004	0.96

Parte B: Modelo Multiplicativo - R^2 intermedio

Parámetros				Estimador	δ			b_{11}			b_{12}			R^2
θ	*ρ*	*M*	σ_2		Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	Coef.	Sesgo	RMSE	
0.0	0.0	-1	1.41	Probit	1.017	0.017	0.003	-2.025	-0.025	0.005	1.018	0.018	0.004	-
				2EMEA	1.019	0.019	0.003	-2.024	-0.024	0.006	1.021	0.021	0.004	0.18
				2EMEM	1.016	0.016	0.004	-2.026	-0.026	0.006	1.020	0.020	0.004	0.33
0.1	0.1	-1	1.41	Probit	1.088	0.088	0.005	-2.124	-0.124	0.008	1.026	0.026	0.004	-
				2EMEA	1.082	0.082	0.005	-2.056	-0.056	0.006	1.021	0.021	0.004	0.18
				2EMEM	1.018	0.018	0.004	-2.030	-0.030	0.006	1.023	0.023	0.004	0.33
0.5	0.6	-1	1.41	Probit	1.501	0.501	0.023	-2.708	-0.708	0.032	1.126	0.126	0.007	-
				2EMEA	1.303	0.303	0.014	-2.051	-0.051	0.007	0.978	-0.022	0.004	0.18
				2EMEM	1.017	0.017	0.004	-2.031	-0.031	0.007	1.021	0.021	0.004	0.33
1.0	0.8	-1	1.41	Probit	2.187	1.187	0.054	-3.637	-1.637	0.074	1.299	0.299	0.015	-
				2EMEA	1.546	0.546	0.025	-1.921	0.079	0.010	0.910	-0.090	0.006	0.18
				2EMEM	1.018	0.018	0.005	-2.040	-0.040	0.008	1.026	0.026	0.005	0.33
2.0	0.9	-1	1.41	Probit	3.615	2.615	0.117	-5.397	-3.397	0.152	1.475	0.475	0.023	-
				2EMEA	2.016	1.016	0.047	-1.634	0.366	0.021	0.780	-0.220	0.011	0.18
				2EMEM	1.025	0.025	0.007	-2.052	-0.052	0.011	1.022	0.022	0.005	0.33

Notar: los valores presentados en las filas rotuladas *probit* corresponden a las estimaciones obtenidas usando el estimador máximo verosímil del modelo probit estándar. A fines comparativos se reescalan esos coeficientes multiplicándolos por el desvío estándar de $u_1 = (1-\rho^2)^{1/2}$.

Tabla 7
Variables independientes- modelo de aprobación de solicitudes para tarjetas de crédito

Income		Ingreso, en diez miles de $
Age		Edad, en años
Curadd	*	Número de años viviendo en la actual dirección
Exp-inc	**	Gasto promedio mensual dividido por el ingreso anual
Avgexp	**	Gasto mensual promedio con tarjeta de crédito
Ownrent	*	Variable binaria que indica propiedad de la casa
Selfempl	*	Variable binaria, =1 si trabajador por cuenta propia
Depndt	*	Número de personas a cargo del solicitante (dependientes)
Inc-per	*	Ingreso por dependiente
Major		Variable binaria, =1 si tiene una tarjeta de crédito principal
Active	*	Número de cuentas de crédito activas
Accounts	*	Número de cuentas abiertas.

Se mantienen los nombres de las variables que emplea Weiss (1999).
* Se incluye sólo en la ecuación de elección binaria. ** Se incluye sólo en la ecuación de *MDR*. Las demás variables se incluyen en ambos modelos.

Tabla 8
Estimación del modelo de aprobación de solicitudes para tarjetas de crédito.

	MV probit estándar	MV modelo de Weiss*	2EMEA**	2EMEM**
Constante	0.496	0.618	1.233	0.555
	(0.196)	(0.175)	(1.069)	(0.231)
Income	0.110	0.089	0.061	0.111
	(0.057)	(0.048)	(0.290)	(0.066)
Age	-0.007	-0.004	0.014	-0.003
	(0.005)	(0.005)	(0.025)	(0.006)
Curadd	0.000	0.000	-0.002	0.000
	(0.001)	(0.001)	(0.002)	(0.001)
Ownrent	0.257	0.194	0.127	0.242
	(0.110)	(0.103)	(0.322)	(0.111)
Selfempl	-0.344	-0.264	0.268	-0.342
	(0.168)	(0.142)	(0.510)	(0.175)
Depndt	-0.134	-0.106	-0.114	-0.126
	(0.073)	(0.064)	(0.167)	(0.078)
Inc-per	0.001	0.01	0.118	0.015
	(0.077)	(0.063)	(0.218)	(0.095)
Major	0.265	0.235	0.573	0.248
	(0.110)	(0.104)	(0.753)	(0.117)
Active	-0.047	-0.032	-0.050	-0.047
	(0.028)	(0.016)	(0.062)	(0.041)
Accounts	0.119	0.087	0.128	0.119
	(0.030)	(0.019)	(0.069)	(0.043)
MDR	-0.944	-1.6	-3.327	-1.704
	(0.146)	(0.099)	(1.199)	(0.413)
MDR2	0.046	0.095	0.024	0.109
	(0.047)	(0.021)	(0.117)	(0.032)
Res. 1° etapa	-	0.887	2.721	0.095
	-	(0.069)	(1.074)	(0.039)
Obs..	1319	1319	1319	1319
Log-Lik.	-486.36	-1485.73	-95.73	-478.14
Pseudo R2	0.307	ND	0.864	0.319
R2 1° etapa	-	-	0.196	0.023

Errores estándar entre paréntesis. * Resultados obtenidos por Weiss (1999). Errores estándar basados en el estimador BHHH/OPG de la matriz de información. ** Errores estándar obtenidos por técnicas de bootstrap (1000 replicaciones). ND: no disponible en el trabajo de Weiss.