

# DETECCIÓN DE INSTRUMENTOS DÉBILES AL CORREGIR ENDOGENEIDAD EN MODELOS LOGIT BINARIOS

Pedro Pablo Navarro Lucic, Universidad de los Andes, Chile pnavarro@miuandes.cl  
C. Angelo Guevara, Universidad de los Andes, Chile, caguevara@miuandes.cl

## RESUMEN

Una variable instrumental es débil cuando no está suficientemente correlacionada con la variable endógena, imposibilitando la corrección de endogeneidad. Este problema ha sido estudiado sólo para modelos lineales. En este trabajo se utilizan simulaciones de Monte Carlo para estudiar la detección de instrumentos débiles en modelos Logit binario, y analizar su impacto en la práctica. Se concluye que, al igual que en modelos lineales, usar instrumentos débiles impide la corrección de endogeneidad en este contexto, y que la debilidad podría detectarse contrastando un test  $F$  con valores críticos muy similares a los que se han sugeridos para modelos lineales.

*Palabras clave: Logit Binario, Endogeneidad, Instrumentos débiles, Funciones de Control.*

## ABSTRACT

An instrumental variable is weak when it is not sufficiently correlated with the endogenous variable, precluding the correction of endogeneity. This problem has been studied only for linear models. In this paper we use Monte Carlo simulations to study the detection of weak instruments in binary Logit models, and to analyze their impact in practice. We conclude that, like in linear models, the use weak instruments precludes the correction of endogeneity in this context, and that weakness could be detected contrasting an F-test with critical values that are very similar to those that have been suggested for linear models.

*Keywords: Binary Logit, Endogeneity, Weak Instruments, Control-function.*

## 1. INTRODUCCION

Un modelo de elección discreta sufre de endogeneidad cuando variables de la parte sistemática de la utilidad están correlacionadas con la parte aleatoria. Esto ocasiona que los estimadores sean inconsistentes.

Por ejemplo, en un modelo de elección residencial los individuos escogen una vivienda considerando variables como precio, número de habitaciones, superficie y calidad de las instalaciones sanitarias. Sin embargo, el investigador que debe construir un modelo que represente el mecanismo de elección, sólo puede observar algunas de esas variables explicativas y debe omitir otras que son difíciles de medir como, por ejemplo, la calidad de las instalaciones sanitarias. Si se omite esta variable, el modelo sufrirá muy probablemente de endogeneidad, debido a que dicha calidad debiera verse reflejada en el precio de la vivienda. El problema es que habrá personas que prefieran pagar más por mayor calidad de instalaciones sanitarias y eso no podrá ser capturado por el modelo. En consecuencia, se confundirán elecciones de viviendas más caras como resultado de la existencia de un parámetro del costo menos negativo de lo que realmente es, en vez de ser el resultado de la disposición a pagar por calidad omitida.

El estado del arte en la corrección de endogeneidad en modelos de elección discreta es el método de funciones de control (Heckman, 1978; Petin y Train, 2002; Guevara y Ben-Akiva, 2006,2012). Este método requiere de variables instrumentales, al igual que el método de mínimos cuadrados en dos etapas para modelos lineales. Las variables instrumentales deben, al mismo tiempo, estar suficientemente correlacionadas con la variable endógena, pero ser independientes del término de error.

Cuando, en modelos lineales, una variable instrumental no está suficientemente correlacionada con la variable endógena, se dice que el instrumento es débil. En dicho caso, los estimadores del modelo lineal corregido son inconsistentes. El impacto y detección de instrumentos débiles sólo ha sido estudiada en la literatura para el caso de modelos lineales, pero no para modelos de elección discreta.

En este trabajo se estudian criterios para detectar instrumentos débiles en modelos de elección discreta del tipo Logit binario. La metodología de este estudio se basa, equivalentemente a lo que se ha hecho en modelos lineales (Stock y Yogo, 2002), en el análisis de la distribución del estadístico  $F$  de la regresión entre la variable endógena y las variables instrumentales. Sin embargo, en este caso, el análisis de los puntos críticos se hace en base a simulaciones de Monte Carlo y no derivando la distribución asintótica. Los resultados se sensibilizan en cuanto al grado de endogeneidad.

Se concluye que, al igual que en modelos lineales, los instrumentos débiles afectan a la corrección de endogeneidad en modelos Logit binario. Además, la evidencia experimental sugiere que un valor del estadístico  $F$  por debajo de 12 garantizaría, con 95% de confianza, que el instrumento sea débil, es decir, que el sesgo del estimador corregido sea mayor a un 10% del sesgo estimador endógeno. Este resultado es similar a lo que se ha recomendado para modelos lineales.

La estructura del artículo es la siguiente. Después de esta introducción se resumen los principales resultados que han sido reportados en la literatura respecto de la detección de instrumentos débiles en modelos lineales utilizando teoría asintótica. Luego, en la Sección 3 se muestra la factibilidad de reproducir los resultados de modelos lineales utilizando simulaciones de Monte Carlo en vez de teoría asintótica. En la Sección 4 se aplica la metodología de simulaciones de Monte Carlo para ilustrar el impacto de la presencia de instrumentos débiles en modelos Logit binario y para establecer criterios para su detección basados en el estadístico  $F$  de la primera etapa del método de funciones de control. Finalmente, en la Sección 5 se resumen los principales resultados y se analizan los alcances y limitaciones de la investigación.

## 2. DETECCIÓN DE INSTRUMENTOS DÉBILES EN MODELOS LINEALES

Considere el modelo lineal descrito en Eq. (1) donde  $X$  es una matriz cuyas columnas corresponden a variables explicativas;  $Y$  es la variable explicada;  $\beta$  es un vector de coeficientes; y  $\varepsilon$  es un vector de errores. Por otro parte, la columna  $X_e$  de  $X$  depende de una matriz  $Z$  de variables instrumentales y un término de error  $\delta$ , como se muestra en la Eq. (2).

$$Y = \beta X + \varepsilon \quad (1)$$

$$X_e = \alpha Z + \delta \quad (2)$$

La endogeneidad en este modelo lineal ocurre cuando los términos de error  $\delta$  y  $\varepsilon$  están correlacionados pues, en dichos caso  $X_e$  estará correlacionado con el término de error  $\varepsilon$ . La endogeneidad en un modelo lineal como el descrito puede corregirse utilizando el método de mínimos cuadrados en dos etapas (2SLS por sus siglas en inglés). El método 2SLS (ver e.g. Wooldridge, 2010) requiere que las variables instrumentales sean exógenas (que no estén correlacionadas con  $\delta$  ni con  $\varepsilon$ ), pero que, al mismo tiempo, sean fuertes (que estén suficientemente correlacionadas con la variable endógena  $X_e$ ).

Cuando los instrumentos están débilmente correlacionados con la variable endógena los estimadores que resultan de aplicar el método 2SLS son inconsistentes. Este problema fue largamente ignorado en la literatura de corrección de endogeneidad en modelos lineales (ver, e.g. Staiger y Stock, 1997). Los primeros trabajos en esta área se concentraron en reportar el impacto del uso de instrumentos débiles. Luego, los estudios se abocaron a la descripción cualitativa de recomendaciones prácticas para detectar la debilidad de los instrumentos. Recién hace aproximadamente una década Stock y Yogo (2002) utilizan teoría asintótica para derivar pruebas estadísticas formales para detectar la debilidad de los instrumentos en la práctica.

El análisis de Stock y Yogo (2002) para la detección instrumentos débiles en modelos lineales, parte por establecer una definición formal de lo que es un instrumento débil en base al concepto de Sesgo Relativo ( $SR$ ) y la cobertura empírica. En este trabajo consideraremos sólo la definición basada en el concepto de  $SR$ , por lo cual éste será analizado con mayor detalle. El  $SR$  se define como la razón entre el sesgo del modelo corregido y el sesgo del modelo sin corregir, tal como se muestra en la Eq.(3)

$$SR = \frac{\hat{\beta}^{corregido} - \beta}{\hat{\beta}^{endogeneo} - \beta} \tag{3}$$

, donde  $\beta$  es el valor real,  $\hat{\beta}^{corregido}$  es el estimador del parámetro corregido obtenido mediante el método de 2SLS, y  $\hat{\beta}^{endogeneo}$  es el estimados del coeficiente del modelo en que no se corrige por endogeneidad.

Stock y Yogo (2002) establecen un límite arbitrario de  $SR > 10\%$  para afirmar que los instrumentos son débiles. Utilizando teoría asintótica los autores muestran que el que  $SR$  sea mayor 10% puede detectarse mediante el análisis del test  $F$  de la regresión de la variable endógena en los instrumentos. Estas recomendaciones se resumen en la Tabla 1, donde se muestra el valor crítico del test  $F$  para el cual la hipótesis nula de que los  $K$  instrumentos son débiles, puede ser rechazada con un 95% de confianza.

**Tabla 1. Recomendaciones por Sesgo Relativo para Detección de Instrumentos Débiles en Modelos de Regresión Lineal**

Sesgo Relativo > 10%	
K	F 95%
3	9.08
5	10.83
10	11.49
15	11.51

Fuente: Stock y Yogo (2002)

El objetivo del presente trabajo es elaborar una tabla de valores críticos equivalente a la Tabla 1, pero para modelos Logit binario. La metodología que se utilizará para este objetivo se basará, no en la teoría asintótica, si no que en simulaciones de Monte Carlo. Con el objeto de validar la metodología propuesta, en la sección siguiente se mostrará que el enfoque se simulaciones de Monte Carlo permite llegar a resultados similares a los de la Tabla 1 para modelos lineales. Luego en la Sección 4 la metodología se aplicará a modelos Logit binario.

### 3. METODOLOGÍA DE SIMULACIONES DE MONTECARLO PARA LA DETECCIÓN DE INSTRUMENTOS DÉBILES: VALIDACIÓN EN MODELOS LINEALES

En esta sección se describe la metodología de simulaciones de Monte Carlo que será utilizada para la definición de límites para detectar instrumentos débiles. Luego, para efectos de validación, la metodología es aplicada a un modelo de regresión lineal y sus conclusiones son comparadas con los resultados teóricos reportados por Stock y Yogo (2002).

La metodología propuesta puede ser descrita en 4 etapas:

- Para un determinado valor de  $\alpha$  en la Eq. (2), número  $K$  de instrumentos, tamaño muestral  $N$ , y grado de correlación  $\rho$  entre  $\varepsilon$  y  $\delta$ , se estiman dos modelos. En el primer modelo no se corrige por

endogeneidad, obteniendo un vector de parámetros  $\hat{\beta}^{endogeneo}$ . En el segundo modelo sí se corrige por endogeneidad, utilizando el método *2SLS*, y se obtienen estimadores  $\hat{\beta}^{corregido}$ .

- Los experimentos se vuelven a generar 5000 veces, pero manteniendo fijos los vectores de variables instrumentales a través de las repeticiones. Esto permite tener un estimador del *SR* para los instrumentos considerados, el cual se calcula como el promedio del *SR* entre las 5000 repeticiones.
- El paso siguiente corresponde a variar la debilidad de los instrumentos, ajustando el valor de  $\alpha$  hasta el punto en que se obtiene un *SR* lo más cercano posible al límite de 10%.
- Finalmente, una vez detectado un valor de  $\alpha$  que se traduce en un *SR* que sea lo más cercano posible al 10%, se procede a estudiar la forma de detectar este fenómeno en la práctica. Esto se logra mediante el análisis del test *F* de la primera etapa del método *2SLS*, de la regresión de la variable endógena en los instrumentos. La generación de 5000 repeticiones permite construir una distribución empírica del test *F*, lo cual hace posible estimar un valor crítico que, con 95% de confianza, permita rechazar la hipótesis nula de que los instrumentos son débiles.

Para validar la metodología propuesta, se comparó el sesgo relativo que se obtenía mediante simulaciones de Monte Carlo, con los sugeridos por Stock y Yogo (2002) para modelos lineales. Con este fin se generó un experimento mediante simulaciones de Monte Carlo, tomando  $N=2000$ ,  $\rho=0.99$ , sobre 5000 iteraciones. Lo que se hizo en este caso fue ajustar el valor del parámetro  $\alpha$  entre 0 y 1 hasta obtener el valor crítico *F* 95%. Para este valor, se tabuló el valor del *SR*, con el fin de verificar si a través de simulaciones, se pueden lograr recomendaciones similares a las propuestas mediante distribuciones asintóticas por Stock y Yogo (2002).

En la Tabla 2 se muestran los resultados de un experimento donde se aproximó el valor crítico *F* 95% mostrado en la Tabla 1, de tal forma de comparar los resultados obtenidos del sesgo relativo (*SR*).

**Tabla 2. Evaluación De Recomendaciones En Modelos Lineales**

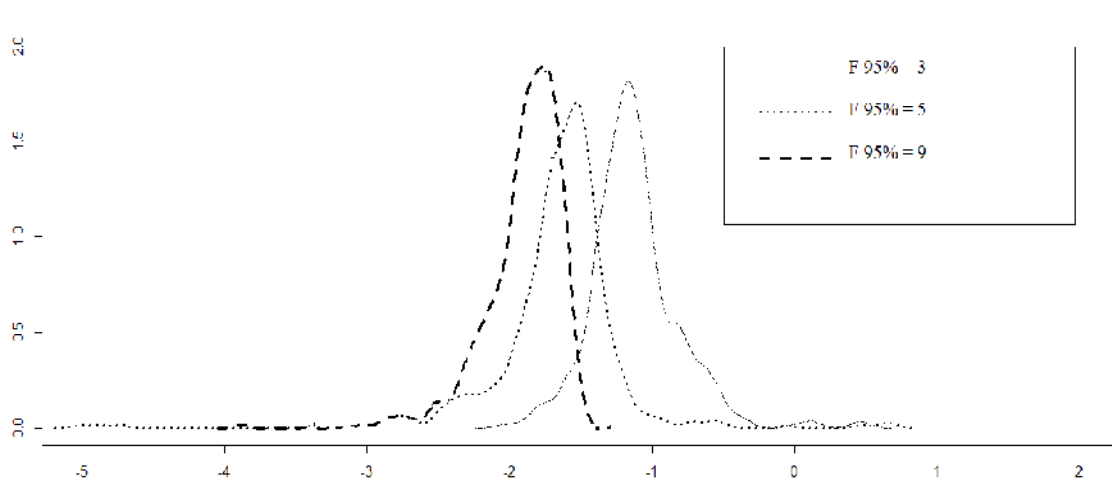
<b>K</b>	$\approx F\ 95\%$	<b>SR</b>
3	9.08	10.03
5	10.83	9.91
10	11.49	10.07
15	11.51	10.09

Los resultados de la Tabla 2 muestran que los valores del *SR* obtenido, tienen un pequeño margen de error respecto al 10%, pero esto se puede comprender, ya que, las estimaciones logradas por Stock y Yogo (2002) son de nivel asintótico, realizadas a través de ecuaciones matemáticas, por tanto más exactas que una aproximación de *F* 95% mediante evaluaciones del parámetro  $\alpha$ .

Adicionalmente, el experimento se repitió para diferentes valores de la correlación  $\rho$ , comprobando que se obtenían también resultados muy similares a los reportados por Stock y Yogo (2002). Estos resultados se muestran en (Navarro, 2013).

Para completar el análisis, en la Figura 1 se muestra la distribución de los estimadores que se obtienen de un modelo corregido para diferentes valores del test *F* de la primera etapa de la

aplicación del método 2SLS. El valor real del estimador de este experimento es -2. El experimento considera 3 instrumentos, 500 iteraciones y un tamaño muestral  $N=500$ .



**Figura 1. Impacto de Instrumentos Débiles en Modelos Lineales. Simulaciones de Monte Carlo**

En la Figura 1 se puede apreciar que mientras se obtenga un mayor valor de  $F$ , la media de la distribución se acerca al valor real -2. Con respecto a la recomendación propuesta en la Tabla 1 para  $K=3$ , podemos ver que efectivamente para valores de  $F$  mayores a 9 la distribución se centra en el valor real con muy poca varianza. Esto refuerza el resultado reportado en la tabla 2, en cuanto a que, al menos para el caso de 3 instrumentos, la metodología de simulaciones entregaría recomendaciones similares a las elaboradas por distribuciones asintóticas.

#### 4. IMPACTO Y DETECCIÓN DE INSTRUMENTOS DÉBILES EN MODELOS LOGIT BINARIO

Para estudiar el impacto y detección de instrumentos débiles en un modelo de elección Logit binario se construyó un modelo de simulación de Monte Carlo. En este caso, la utilidad de cada una de las dos alternativas se modeló en base a lo que se indica en la ecuación (4).

$$U_{in} = -2p_m + 1a_m + 2b_m + \varepsilon_m \tag{4}$$

, donde  $U_{in}$  representa la utilidad de la observación  $n$  para la alternativa  $i$ ;  $p$ ,  $a$  y  $b$  son atributos de las alternativas y  $\varepsilon$  es un término de error. Los atributos y el término de error fueron generados de la siguiente forma:

- $a_i, b_i$ , vectores independientes e idénticamente distribuidos Unifome (-3,3)
- $\varepsilon_i, \delta_i$ , vectores Normal (0,1) con cierto nivel de correlación  $\rho$
- $z_{ik}$ ,  $K$  vectores independientes e idénticamente distribuidos Normal (0,1)

$$p_{in} = -5 + \sum_{k=1}^K \alpha Z_{ik} + \delta_{in}, \quad (5)$$

donde  $\alpha$  es un escalar entre 0 y 1.

Dado que los errores,  $\varepsilon$  y  $\delta$  están correlacionados ( $\rho \neq 0$ ), el modelo sufre de endogeneidad, pues la variable  $p$  estará correlacionada con el error  $\varepsilon$  en la ecuación estructural (4). Esto significa que la estimación usual mediante máxima verosimilitud resultaría en estimadores inconsistentes. Para salvar esta limitación es posible aplicar el método de funciones de control.

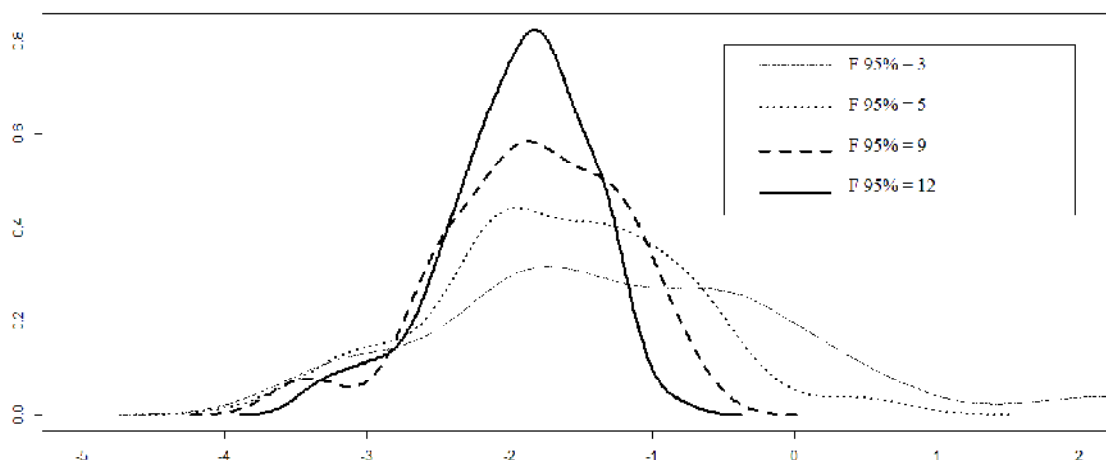
La corrección de endogeneidad mediante el método de funciones de control en modelos Logit consiste de dos etapas (ver, por ejemplo, Petin y Train, 2002; Guevara y Ben-Akiva, 2006,2012). En la primera etapa se realiza una regresión de mínimos cuadrados ordinarios de la variable endógena en las variables instrumentales apropiadas. En la segunda etapa se utilizan los residuos de la estimación realizada en la primera etapa como variables auxiliares que se agregan a la ecuación estructural del modelo. Cabe destacar que la estimación simultánea del método de funciones de control es, en general, más eficiente que la estimación en dos etapas, pero que ambas metodologías son consistentes (Guevara y Ben-Akiva, 2012). En este artículo el análisis se realiza íntegramente mediante estimación secuencial debido a que la metodología propuesta requiere hacer miles de estimaciones, lo cual se vuelve poco práctico con estimación simultánea.

Por la forma en que fueron construidos los vectores  $z_{ik}$ , éstos son instrumentos apropiados para la variable endógena  $p$ . Por una parte los instrumentos  $z_{ik}$  son generados exógenamente al modelo y, por otro lado, debido a la ecuación (5), ellos están correlacionados con la variable endógena  $p$ . El nivel de correlación con  $p$ , y por lo tanto del poder del instrumento, dependerá del parámetro de modelación  $\alpha$ .

Cabe destacar que, dado el término de error  $\varepsilon$  es normal en la ecuación (4), el modelo de elección descrito es formalmente un modelo Probit binario. Sin embargo, para simplificar el proceso de estimación de las repeticiones, el modelo fue estimado como un Logit binario. De acuerdo a Lee (1982), Ruud (1983) y Cramer (2007) este tipo de error de especificación no es severo, en el sentido de no comprometer la posibilidad de obtener estimadores consistentes hasta una escala.

Partiremos utilizando este modelo de Monte Carlo para estudiar el impacto de la utilización de instrumentos débiles en la corrección de endogeneidad en modelos de elección Logit binario. Los resultados del análisis del impacto de la debilidad de los instrumentos son resumidos en la Figura 2, donde se muestra la distribución de la razón entre los parámetros de la variable  $p$  y de la variable  $\alpha$ . Cabe destacar que es necesario en este caso estudiar la razón de los estimadores pues, en este caso, la corrección de endogeneidad implica potencialmente un cambio de escala (Guevara y Ben-Akiva, 2012)

El experimento utilizado para generar la Figura 2 considera que  $K=3$ ,  $N=500$  y 500 repeticiones. Variando el valor de  $\alpha$  se generaron experimentos en los el valor del estadístico  $F$  95% de la primera etapa del método de funciones de control tomaba diferentes valores. Los valores elegidos para el test  $F$  son los mismos que los reportados en la Figura 1 para modelos lineales (3, 5 y 9) más el valor  $F$  95% = 12.



**Figura 2. Consecuencia de Instrumentos Débiles en Modelo Logit Binario**

En la Figura 2 puede verse que para  $F\ 95\%=12$  la corrección de la endogeneidad es exitosa, pues la distribución muestral de la razón de los estimadores está centrada en el valor real  $-2$  y tiene colas relativamente simétricas. En cambio, a medida que el valor de  $F$  disminuye, la calidad de la corrección empeora, llegando incluso a observarse distribuciones bi-modales.

Un aspecto que es notoriamente diferente entre la corrección de endogeneidad en modelos lineales (Figura 1) y discretos (Figura 2) es que, en el primer caso, a medida que el valor de  $F$  disminuye, la media muestral cambia, pero la varianza muestral se mantiene. En cambio, en el caso de los modelos de elección discreta, al disminuir  $F$ , la varianza muestral aumenta de manera clara. Es posible que esta diferencia entre modelos lineales y discretos se explique por el cambio de escala que ocurre en los segundos al corregir por endogeneidad (ver Guevara y Ben-Akiva, 2012). Otra posible explicación puede estar relacionada con el uso de la versión secuencial del método de funciones de control, en vez de utilizar un enfoque simultáneo. Del mismo modo, es plausible que este comportamiento sea el resultado de haber generado los errores  $\varepsilon$  y  $\delta$  multivariados Normal, en vez de hacer que  $\varepsilon$  fuese valor extremo. El estudio detallado de este fenómeno es dejado para análisis futuro.

El segundo paso en la investigación corresponde a verificar si es posible identificar valores críticos del test  $F$  de la primera etapa del método de funciones de control que permitan detectar la debilidad de las variables instrumentales.

La metodología para encontrar los valores críticos fue la siguiente. Para un experimento dado, se varió el valor del parámetro  $\alpha$  hasta el punto en que el sesgo relativo promedio de las repeticiones fuese lo más cercano posible a 10%. Luego, se identificó en valor 95% de la distribución del test  $F$  a lo largo de las repeticiones. El experimento consideró un tamaño muestral de 2000 y 5000 repeticiones. Además, se analizó el impacto de variar el nivel de endogeneidad ajustando el parámetro  $\rho$ . Se consideraron tres valores de  $\rho$   $\{0.25, 0.5, 0.75\}$ .



La Tabla 3 muestra los resultados obtenidos al estimar el problema con un Modelo Logit, para distintos valores de  $\rho$ . La columna 1, muestra el número de instrumentos (K), la columna 2, muestra el sesgo relativo  $SR$  que se alcanzó (aproximadamente 10%), y la columna 3 muestra los valores críticos de  $F$  95%.

**Tabla 3. Valores Críticos para Detección de Instrumentos Débiles en Modelo Logit Binario.**

K	$\rho=0.25$		$\rho=0.5$		$\rho=0.75$	
	$SR \approx 10\%$	F 95%	$SR \approx 10\%$	F 95%	$SR \approx 10\%$	F 95%
3	10	13.42	10	10.77	10.06	9.46
5	10.05	13.15	10.03	11.57	10.06	10.53
10	10.03	13.34	9.99	12.07	10.02	11.37
15	10	12.21	10.06	11.98	10.05	11.62

Se observa que el valor obtenido  $F$  95% va disminuyendo levemente a medida que aumenta la endogeneidad (mayor  $\rho$ ) y, al mismo tiempo, se acerca a los valores recomendados en modelos lineales. En términos numéricos se puede decir que  $F$  95% tiende a un valor entre 11 y 12 a medida que aumentan los instrumentos, que es justamente lo que proponen Stock y Yogo (2002) para modelos lineales mediante el análisis de distribuciones asintóticas.

## 5. CONCLUSIONES

En este trabajo se utilizan simulaciones de Monte Carlo para estudiar la detección de instrumentos débiles en modelos Logit binario, y analizar su impacto en la práctica.

Como primer resultado se concluyó que, al igual que en modelos lineales, los modelos Logit Binarios también se ven afectados en presencia de instrumentos débiles al corregir por endogeneidad (Figura 2). Sin embargo, se observa una diferencia con los modelos lineales, en cuanto a que la distribución del estimador resulta con una mayor varianza que en modelos lineales. El análisis de la naturaleza e implicancias de esta diferencia se deja para estudios futuros.

Por otro lado se desarrolló una metodología basada en simulaciones de Monte Carlo que permitiera identificar valores críticos para detectar instrumentos débiles. Esta metodología fue primero validada en modelos lineales y luego aplicada a modelos de elección Logit binario, donde se mostró que los valores críticos en este caso son muy similares a los reportados previamente para modelos lineales.

En términos prácticos, la evidencia desarrollada en esta investigación sugiere que valores del test  $F$  de la primera regresión del método de funciones de control, que sean inferiores a 12, permitirían afirmar, con un 95% de confianza, que los instrumentos son débiles.

Se identifican cuatro posibles extensiones directas a este trabajo. El primer paso corresponde a aplicar la metodología de Monte Carlo utilizada en este estudio, pero considerando la definición

alternativa de instrumento débil sugerida por Stock y Yogo (2002), la cual se basa en la cobertura empírica. Una segunda línea de análisis corresponde a la variación del sesgo relativo límite de (10%), a valores más pequeños (5%) y mayores (15%). Tercero, se propone estudiar el impacto de la varianza relativa del término de error, respecto de las variables observadas y el impacto de utilizar una aproximación Normal multivariada, en vez de una distribución multivariada valor extremo para generar los experimentos. Por último, es necesario investigar la naturaleza e implicancias del notorio cambio en la varianza muestral al variar el poder de los instrumentos en modelos binarios.

### **Agradecimientos**

Esta publicación fue financiada en parte por CONICYT, FONDECYT/Iniciación 11110131. Todos experimentos fueron generados y/o estimados utilizando el software open-source R (R Development Core Team, 2008).

### **Referencias**

Cramer, J. (2007), "Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances," **Oxford Bulletin of Economics and Statistics**, 69(4), 545-555.  
**Econometrica** , 65, 557-586

Guevara, C. A., & Ben-Akiva, M. E. (2012). "Change of Scale and Forecasting with the Control-Function Method in Logit Models". **Transportation Science**, 46(3), 425-437.

Guevara, C. and M. Ben-Akiva (2006), "Endogeneity in Residential Location Choice Models," **Transportation Research Record**, 1977, 60-66.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," **Econometrica**, 46, 931-959.

Lee, L. (1982), "Specification Error in Multinomial Logit Models," **Journal of Econometrics**, 20, 197-209.

Navarro, P. P. (2013), "Detección de Instrumentos Débiles al Corregir Endogeneidad en Modelos Logit", **Memoria de Título de Ingeniería Civil**, Universidad de los Andes, Chile.

Petrin, A. and K. Train (2002), "Omitted Product Attributes in Discrete Choice Models," **Working Paper, Department of Economics, University of California, Berkeley, CA.**

R Development Core Team (2008), R: A Language and Environment for Statistical Computing, **R Foundation for Statistical Computing**, Vienna, Austria, <http://www.Rproject.org>.

Ruud, P. (1983), "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Models," **Econometrica**, 51, 225-228.

---

Staiger, D. y Stock, J. (1997). “Instrumental Variables Regression with Weak Instruments.”

Stock, J. H., and Yogo, M. (2002), “Testing for Weak Instruments in Linear IV Regression,” **NBER Technical Working Paper** No. 284

Wooldridge, J. (2002), **Econometric Analysis of Cross-Section and Panel Data**, MIT Press, Cambridge, MA.