

Endogeneidad en modelo de determinantes de la pobreza.

Tabare Fernandez.

Cita:

Tabare Fernandez (2017). *Endogeneidad en modelo de determinantes de la pobreza. XXXI Congreso de la Asociación Latinoamericana de Sociología. Asociación Latinoamericana de Sociología, Montevideo.*

Dirección estable: <https://www.aacademica.org/000-018/1051>

Endogeneidad en un modelo de determinantes de la pobreza multidimensional

Dr. Tabaré Fernández Aguerre
tabare.fernandez@cienciassociales.edu.uy
Departamento de Sociología,
Facultad de Ciencias Sociales
Universidad de la República
Rivera, noviembre de 2017

RESUMEN

El objetivo de este trabajo es presentar el análisis hecho del supuesto de exogeneidad y su correlato, la endogeneidad, en un modelo de regresión logística binaria sobre determinantes de la pobreza en Uruguay. El problema que enfrentó la investigación puede resumirse de la siguiente forma. Las explicaciones de la pobreza que han sido desarrolladas por lo general tienen por variable dependiente, una medida monetaria. Existe, en términos relativos, un muy reducido desarrollo del estudio de determinantes que ocupe medidas multidimensionales. Esto conlleva que cuando se quiere testear la hipótesis de la estabilidad de la estructura de determinantes en uno y otro caso, se encuentre que algunos conceptos que están expresados como variables del lado derecho de la ecuación, también han sido tomados en el lado izquierdo, en el vector (de logros o funcionamientos) que permite identificar la pobreza multidimensional. En presencia de endogeneidad los estimadores aplicados ya no cumplen con los estándares estadísticos básicos de ser los mejores, con menor varianza, insesgados e inconsistentes. La estimación del efecto de una variable regresora endógena en un modelo no corregido será siempre inconsistente, es decir diferente a su efecto real, por más que se aumente indefinidamente la cantidad de observaciones con las que se hace la estimación. Las evidencias halladas a través de distintas pruebas justificaron la aplicación de métodos para una prueba más selectiva y rigurosa del problema. En la bibliografía se encuentran en forma estándar varios métodos cuya aplicación depende del problema identificado en general, así como de la distribución de la variable endógena, entre los cuales se encuentran: i) variables instrumentales; ii) la estimación máximo verosímil; iii) biprobit y iv) la función de control ("control function" o CF). Finalmente, luego de varios exámenes alternativos, llegamos a la conclusión de que no existen métodos consensuados para corregir este problema, contemplando todos los supuestos que están en juego.

ABSTRACT

The objective of this paper is to present the analysis made of the assumption of exogeneity and its correlate, endogeneity, in a model of binary logistic regression on determinants of poverty in Uruguay. The problem faced by the investigation can be summed up as follows. The explanations of poverty that have been developed usually have a dependent variable, a monetary measure. There is, in relative terms, a very small development of the study of determinants that occupies multidimensional measures. This implies that when you want to test the hypothesis of the stability of the structure of determinants in one and another case, it is found that some concepts that are expressed as variables on the right side of the equation, have also been taken on the side Left, in the vector (of achievements or performances) that allows to identify the multidimensional poverty. In the presence of endogeneity, the applied estimators no longer meet the basic statistical standards of being the best, with the lowest variance, unbiased and inconsistent. The estimation of the effect of an endogenous return variable in an uncorrected model will always be inconsistent, that is to say different from its actual effect, however the number of observations with which the estimation is made is indefinitely increased. Evidence found through various tests justified the application of methods for a more selective and rigorous test of the problem. In the bibliography, several methods whose application depends on the problem identified in general, as well as the distribution of the endogenous variable are found in standard form, among which are: i) instrumental variables; (ii) the maximum likelihood plausible estimate; (iii) Biprobit and IV) the control function ("control function" or CF). Finally, after several alternative examinations, we concluded that there are no agreed methods to correct this problem, contemplating all the assumptions that are at stake.

Palabras clave

(Incluir 3 palabras clave en español o portugués)

Keywords

(Incluir 3 palabras clave en inglés)

Indice

Presentación y objetivos	5
Planteo del supuesto de exogeneidad	6
Identificación de la endogeneidad	7
Prueba de endogeneidad y “variables instrumentos”	8
Limitaciones y supuestos del enfoque de la función de control	9
Determinantes de la pobreza multidimensional: modelo bajo sospecha.	11
Endogeneidad teórica de los regresores	14
Primer ejercicio: prueba estadística aplicada al modelo de determinantes	15
Segundo ejercicio: estimación por CF	17
Discusión de resultados y conclusiones	19
Referencias	21

Presentación y objetivos¹

La motivación de este trabajo se encuentra en dos trabajos de investigación recientemente concluidos, uno en el ámbito de los determinantes de la pobreza y otro en el estudio de las transiciones de la educación al trabajo (Fernández, Borrás, & Ezquerro, 2017; Fernández, Marques, & Lorenzo, 2017). En ambos temas, colegas y revisores nos sugirieron un examen de endogeneidad de los modelos ajustados. En ambos casos, y este en parte enmarca este trabajo, se trataba de modelos con una variable dependiente dicotómica.

Motivado en esta observación, he revisado una diversidad de textos de referencia en la enseñanza de la econometría nivel intermedio y avanzado, tales como el de William Green, Jeffrey Wooldridge, y el de Colin Cameron y Pravin Trivedi (Cameron & Trivedi, 2010; Wooldridge, 2010; Greene, 2012). Así mismo, hallé varias presentaciones, discusiones en workshops, congresos, artículos en revistas estadísticas, preguntas y respuestas en listas de FAQ de estadística (v.g. stataлист). Las preguntas más específicas variaban desde casos más o menos “típicos” de cátedra, ejercicios de simulación, cuestiones de salud en personas retiradas, migración internacional, elecciones educativas, etc (Baum, Dong, Lewbel, & Yang, 2012).

Es de notar que los debates son al punto de importantes que ocupan cada vez más espacio en un texto clásico como el de William Greene, si uno ve la quinta edición (2002) y la séptima edición (2012) (Greene, 2012). El orden en que se presenta este tema también ha variado en la exposición de los textos: ya no es algo especial a ser tratado una vez agotado el modelo normal (es decir, varios cientos de páginas desde los comienzos) (Gujarati, 2004), sino que el problema de los regresores endógenos impulsa a presentar conjuntamente y desde los primeros capítulos el modelo lineal de regresión clásico (es decir, ajustado por Ordinary Least Squares, OLS) conjuntamente con el modelo de regresión en dos etapas (Two stages least square, 2SLS) (Wooldridge, 2010).

La creciente atención que ha ido recibiendo en la bibliografía estadística y empírica, da la impresión que está generada por el creciente uso de modelos econométricos para testear hipótesis causales, entre ellas aquellas relativas a los efectos de ciertos tratamientos implementados por políticas públicas como puede ser un programa de lucha contra la pobreza, el impacto de un evento de vida en un cambio de comportamiento en la salud como puede ser la jubilación, los efectos que tiene la educación técnica sobre una inserción exitosa de los jóvenes en el mercado de trabajo, el impacto de un tratamiento médico cuyos beneficiarios son seleccionados de forma no aleatoria, etc. Es decir, el tema emerge por diversos caminos, todos ellos vinculados con (al menos) una variable independiente incluida en un modelo de regresión de interés (llamado por algunos “estructural” o “principal”) cuyos valores se sospechan están condicionados por otras variables independientes, alguna de ellas también incluidas en el modelo de regresión “principal” (son un subconjunto de los regresores especificados) y otras que no lo están; esto es han sido omitidas. En síntesis, podría decirse que el examen de la presencia de endogeneidad (o más correctamente de acuerdo al proceder estándar de la estadística, el rechazo del supuesto de exogeneidad de los regresores) constituye un caso particular dentro del capítulo general del estudio de la especificación del modelo de regresión. Un modelo con endogeneidad es un modelo mal especificado cuyos estimados serán inconsistentes.

¹ Deseo agradecer en forma expresa, la atención y orientación que me brindara el dr. Carlos Casacuberta sobre el problema de endogeneidad hallado en el primero de los trabajos. También agradezco algunos comentarios y sugerencias que me diera Cecilia Rodríguez respecto a los resultados finales. Obviamente, esta exposición y las soluciones discutidas son de responsabilidad del autor.

Planteo del supuesto de exogeneidad

Por lo general, la bibliografía más antigua trataba el tema de exogeneidad como una parte más general del problema de la inferencia causal en modelos econométricos, aunque todos los autores reconocen la complejidad del tema, en particular delicado cuando el objetivo es hacer inferencias de tipo causal. Sin embargo, tal como contemporáneamente se entiende, el problema no tiene que ver solo con la inferencia (v.g. pruebas de hipótesis), sino también con la misma estimación.

En todos los casos se parte de un modelo de regresión principal, cuyo propósito es explicar la distribución condicional de la variable dependiente “y”, dado un número J de regresores “x”, tal que $j=1,2,3,\dots, J$. Expresaré el modelo estructural en términos lineales simplemente por comodidad para esta primera etapa.

$$[1] y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_j x_j + \varepsilon$$

El modelo lineal tiene por supuesto que los regresores sean exógenos en un sentido fuerte, esto es:

$$[2] \text{Cov}(x_j, \varepsilon) = 0 \quad \forall j = 1, 2, \dots, J$$

$$[3] E(\varepsilon/x_1, x_2, x_3 \dots x_j) = E(\varepsilon/X) = 0$$

Lo que enuncia la ecuación 2 es que para cualquier regresor j-ésimo incluido en un modelo de regresión, la covariación con el término de error es igual a cero. Simplificando rústicamente el asunto, todas las variables descartadas del modelo principal al término de error no tiene relevancia *tampoco* para explicar variaciones en los regresores. Para una más ágil notación, expondré la ecuación [1] en algebra matricial, donde \mathbf{X}' es un vector $1 \times J$ de variables que incluye la unidad como uno de sus elementos.

$$[4] y_1 = \mathbf{X}'\beta + \varepsilon$$

$$[5] E(\mathbf{X}'\varepsilon) = 0$$

La violación de este supuesto se llama “endogeneidad”. Esto ocurre cuando el término de error “contiene” información correlacionada con alguno de los regresores. Esta violación puede ocurrir con uno, dos o con varios regresores del modelo principal. Solo por simplicidad, por el momento, supondré que sólo uno de estos regresores está afectado por este problema. Del lado derecho de la ecuación habrá un vector $1 \times K$ de regresores exógenos², \mathbf{X}'_1 , y un regresor endógeno, y_2 . Para distinguirlo de los demás efectos, α_1 es el estimado del efecto marginal del regresor endógeno y_2 . Denominaré a este modelo como el modelo original.

$$[6] y_1 = \mathbf{X}'_1 \gamma + \alpha_1 y_2 + \varepsilon$$

² Estrictamente, \mathbf{X}'_1 se define como un subvector de \mathbf{X}' que incluye a la constante, por lo que se cumple con que $J - K \geq 1$.

En particular, debe tenerse la mayor cautela cuando el estudio se focaliza precisamente en una evaluación del efecto marginal de γ_2 , por ejemplo, cuando esta mide un tratamiento de política pública. La estimación del efecto de la variable regresora endógena a la que se atribuye efecto causal, será siempre inconsistente, es decir diferente a su efecto real, por más que se aumente indefinidamente la cantidad de observaciones con las que se hace la estimación. Esto es, aún en muestras grandes, el estimado del efecto marginal no convergerá en el parámetro poblacional. En síntesis:

$$[7] \hat{\alpha}_1 \neq \alpha_1$$

Ahora bien, conviene subrayar que no solo está sesgada la estimación del efecto marginal el regresor endógeno. En presencia de endogeneidad *todos los* estimadores (β, α) ya no cumplen con los estándares estadísticos básicos de ser los mejores, con menor varianza, insesgados e inconsistentes. Categóricamente ha de decirse que la presencia de endogeneidad invalida la información tanto sobre la significación del coeficiente (esto es, la prueba de independencia estadística) como el signo o la magnitud del efecto marginal.

La bibliografía revisada consensua en que el supuesto de exogeneidad tiene una utilidad *limitada* en varios campos temáticos. Incluso, se señala que plantearlos como problemas (a evitar) puede resultar debatible en la medida en que desde la teoría misma se espera observar, por ejemplo, simultaneidad entre variables. Por tanto, podríamos sostener que, en las ciencias sociales en general, resulta más razonable esperar la violación de este supuesto en forma frecuente y para una amplia diversidad temática³.

Los dos campos temáticos que motivaron la realización de este análisis son precisamente típicos. Varios de los determinantes incluidos en la explicación de la condición de pobreza de un hogar por lo general son endógenos: la titularidad con que un hogar ocupa una vivienda suele ser resultado sea de diversos mecanismos que tienen que ver con la herencia, el mercado de trabajo, el capital humano o el área geográfica, entre otros. A su vez, cuando se estudia el impacto de la educación técnica sobre la inserción laboral resulta imprescindible analizar cómo una persona eligió cursar esa modalidad. Esto es, la opción por la modalidad Técnica en la Educación Media está relacionada tanto con la existencia local de oferta de esta modalidad como con variables como el género, la ocupación de los padres, el costo de oportunidad y la trayectoria educativa. Algunas de estas variables suelen estar también en la ecuación principal que explica la inserción laboral pero otros no son considerados por no tener aparente relación con este resultado.

Identificación de la endogeneidad

Como en varios otros temas econométricos, una vez definido en términos generales es necesario pasar a tratar sucesivamente los problemas de *identificación*, *prueba* y *corrección del problema* (Gujarati, 2004). Este protocolo sencillo permitirá a continuación mostrar algunas de las complejidades y problemas que tiene la endogeneidad.

La lectura de la bibliografía deja en primer lugar una conclusión simple pero nada trivial: la *identificación* de la endogeneidad es antes que nada teórica. Esta situación puede originarse por varias razones. Greene (2012: 219-222) sistematiza 8 casos típicos: i) omisión de variables; ii) tratamientos endógenos; iii) ecuaciones simultáneas; iv) modelos dinámicos de panel; v) omisión de parámetros de heterogeneidad no aleatorios en la comparación entre países o regiones por ejemplo; vi) error de medición; vii) muestreo no aleatorio; viii) "attrition" o "pérdida" de casos en los estudios de panel

³ Tal vez valga decir que esta era la posición que Greene sostenía en la tercera edición de su libro sobre Análisis Económico a principios de los años noventa.

Sólo en pocas ocasiones, la distinción entre las diversas formas de endogeneidad es nítida. Un modelo puede estar afectado a la vez por la omisión de variables relevantes (tipo i) y existir tratamientos endógenos (tipo ii).

La forma funcional del modelo y la temporalidad adicionan problemas. Sea por razones teóricas o por razones metodológicas (en la especificación), los modelos lineales en su forma más simple ignoran el momento o la forma en que se genera cada variable, esto es, la matriz de información \mathbf{X} , tiene la misma temporalidad cuando esto no es teórica ni empíricamente ajustado. A su vez, la forma aditiva de especificación del modelo lineal ignora la interacción recíproca entre regresores, o sea, la común generación de algunos de ellos por parte de otros factores que permanecen latentes. Estos a su vez, puede ser el caso de que generen tanto la variable dependiente como (al menos) una variable independiente.

Prueba de endogeneidad y “variables instrumentos”

Un segundo paso importante en el estudio del supuesto de exogeneidad consiste en la realización de una prueba estadística. Este tipo de pruebas tiene la característica de ser una prueba de especificación. En términos muy generales, se procede a especificar y estimar un modelo donde *no se viola el supuesto* porque se han incluido estos regresores inicialmente excluidos (denominamos este modelo como completo) y se comparan los estimados con aquel modelo que se sospecha que tiene endogeneidad (modelo principal). Esto es, el procedimiento implica partir del modelo presentado más arriba:

$$[8] \quad y_1 = \mathbf{X}'_1 \gamma + \alpha_1 y_2 + \varepsilon_1$$

Especificar una ecuación nueva, denominada “ecuación reducida”, en la que se cumple estrictamente el supuesto de exogeneidad:

$$[9] \quad y_2 = \mathbf{X}'_2 \delta + u_2$$

$$[10] \quad E(\mathbf{X}'_2 u_2) = \mathbf{0}$$

Resulta importante observar que en la ecuación anterior se ha incluido otro subvector de \mathbf{X}' , con $1 \times L$ regresores y designado como \mathbf{X}'_2 estrictamente debe contener *al menos un* regresor que no esté incluido en \mathbf{X}'_1 , el cual es designado como “variable instrumental”⁴. De las ecuaciones [8] y [9] surge que la endogeneidad de y_2 emerge como consecuencia sólo si ε_1 está correlacionado con u_2 . Expuesta esta noción en términos lineales:

$$[11] \quad \varepsilon_1 = \rho_1 u_2 + u_1$$

Esto permite incluir [1] en la ecuación [6] y re-escribir la expresión en los siguientes términos:

⁴ Cameron & Trivedi (2010:183) dividen la exposición en esta materia entre los modelos donde existe el mismo número de regresores endógenos que variables instrumentales, y aquellos en que hay más variables instrumentales que regresores endógenos. Estos últimos son denominados “sobreidentificados”.

$$[12] \quad y_1 = \mathbf{X}'_1 \gamma + \eta_1 y_2 + \rho_1 v_2 + u_1$$

Esta expresión nueva corrige la omisión de variables incluyendo como regresor el residuo de la ecuación reducida [8] como uno de los determinantes de la variable dependiente de interés. DE la ecuación [11] se obtiene una prueba directa sobre la hipótesis de exogeneidad. Si el regresor y_2 es realmente exógeno, entonces no será posible descartar la hipótesis nula sobre el coeficiente asociado al término “un”, esto es: $\rho_1 = 0$. Esta prueba es denominada por la bibliografía como Durbin-Wu-Hausman (DWH).

Existe otra posibilidad también de probar la hipótesis nula. Parte de la base de que este procedimiento generará un vector de coeficientes, γ , eficientes y consistentes (el modelo estructural) y deberá ser comparado con el vector de coeficientes estimados en el modelo con variable omitida, β . Esta lógica guía el principio de la prueba de Hausman sobre especificación. La hipótesis nula sostiene que ambos vectores son estadísticamente iguales, esto es, consistentes⁵. El rechazo de la hipótesis nula implicará que los coeficientes del modelo principal son inconsistentes.

$$[13] \quad H_0: \beta = \gamma; \quad \alpha = \eta \quad ,$$

El test de Hausman para el caso de un único regresor endógeno, donde \hat{V} es la matriz de varianzas y el estadístico:

$$[14] \quad T_H = \frac{(\eta_1 - \alpha_2)^2}{\hat{V}(\eta_1 - \alpha_2)}$$

Donde $T_H \sim \chi^2$ con un grado de libertad.

Limitaciones y supuestos del enfoque de la función de control

Llegado a este punto de la argumentación es necesario indicar que el supuesto que me he basado para el desarrollo es muy restrictivo: solo abordé una situación ideal en la que la variable dependiente sigue una distribución (aproximadamente) normal, sólo hay una variable endógena y esta también sigue una distribución (aproximadamente) normal. Este es el caso ideal para el que se desarrolló el procedimiento descrito en la sección anterior, el cual se denomina “enfoque de la función de control” (CF) (Wooldridge, 2010) o también “enfoque de modelo estructural” (Cameron & Trivedi, 2010). Es bastante robusto, y permite llegar a las mismas conclusiones que con el método de variables instrumentales de Heckmann, siempre que, de nuevo, se cumpla la condición de existir una única variable endógena de tipo métrica. Ha sido extensamente utilizado a partir de las recomendaciones que Wooldridge (2010) hace en su texto de econometría, pero tiene varias limitaciones y restricciones que conviene explicitar a esta altura de la exposición, aun dentro del muy especial caso ideal que estoy presuponiendo.

⁵ En tal caso, los estimados por el método de 2SLS serán empero, ineficientes.

La primera limitación que resulta de importancia, es que el enfoque CF requiere que y_2 sea una variable métrica. Adicionalmente, requiere que y_1 sean también métrica y que por lo tanto el ajuste de las ecuaciones pueda hacerse mediante el método de mínimos cuadrados ordinarios (OLS, en inglés). Pero es de señalarse que aún sin cumplirse esto último, el procedimiento CF puede ser extendido a modelos no lineales, aunque habrá de toparse con un problema importante que será tratado más abajo.

Una segunda limitación resulta de la cantidad de variables endógenas supuestas en el modelo original: el enfoque CF tiene resultados destacados cuando hay un único regresor endógeno. No está claro cuál puede ser la consecuencia de violar este supuesto y enfrentar un modelo con más de un regresor endógeno. ¿Habría que aplicar tantas funciones CF como variables endógenas haya?.

Además, existen supuestos que deben cumplirse al seguir con este enfoque. En primer lugar, debe ser posible especificar de forma correcta un modelo explicativo para la variable hipotéticamente endógena. Esto se logra con un vector de regresores relacionados con y_2 , siempre que estos regresores no estén relacionados con y_1 . Esto es, se debe contar con “buenos instrumentos” y obviar el problema de los “instrumentos débiles”.

En segundo lugar, el enfoque CF expuesto más arriba requiere que se verifique el supuesto de homocedasticidad. Evidentemente, esto no se cumple en el caso de que la variable dependiente, y_1 , sea dicotómica. Pero no es el único caso en que podría ocurrir esto. Aún así la bibliografía consultada sugiere ajustar el modelo estructural utilizando el método de 2SLS (por two stage least square) con previsión de heterocedasticidad. Las estimaciones aún serían robustas en este caso.

Sin embargo, existen pocos elementos informativos sobre las consecuencias de aplicar estos métodos en el caso de que: a) la variable dependiente sea dicotómica o que b) la variable endógena sea dicotómica. Para entender el origen de las complejidades es preciso ampliar el campo de problemas y distinguir cuatro tipos, aunque por el momento sea solo combinando dos modelos de distribución posibles, la normal (N) y la binomial, B, para las variables implicadas, tal como se presentan en el esquema n^o2.

Esquema n^o1. Modelo de distribución de la variable dependiente y variable regresora endógena.

	Variable $y_1 \sim N$	Variable $y_1 \sim B$
Variable $y_2 \sim N$	I	II
Variable $y_2 \sim B$	III	IV

El tipo de situaciones más estudiado y que dispone de soluciones robustas es aquel en el que tanto la variable dependiente del modelo principal, y_1 , como aquella regresora sospechosa por endogeneidad, y_2 , son métricas con una distribución normal (o al menos relativamente simétrica, sin colas pronunciadas ni una elevada curtosis). Este tipo de situaciones permite aplicar el método de mínimos cuadrados (OLS) tanto en las pruebas como en corrección del problema. Salvo en casos donde existe sobreidentificación de instrumentos o un número relativamente alto de instrumentos en relación al tamaño de la muestra, entonces una solución del tipo 2SLS será la más apropiada. Es evidente que dos problemas que motivaron este estudio no están dentro de este tipo de situaciones. Y si se me permite, se trata de problemas bastante frecuentes en los campos de la sociología.

Precisamente, el tipo IV es el opuesto al anterior, sea en términos de prueba sea en términos de corrección. La variable dependiente sigue una distribución binomial y la regresora endógena también. Una solución de tipo 2SLS podría ser utilizada, incluso asumiendo y corrigiendo por heterocedasticidad, pero existe un escepticismo en la bibliografía de si es posible contar con pruebas robustas para examinar el supuesto, y más aún, si la aplicación de alguna de las alternativas de corrección realmente entrega estima-

ciones consistentes (Dong & Lewbel, 2012; Dong & Lewbel, 2015; Cameron & Trivedi, 2010; Baum, Dong, Lewbel, & Yang, 2012). Existe consenso en que la ecuación [12], provee de una prueba aceptable sobre la endogeneidad de y_2 pero no provee estimadores consistentes cuando no se cumple el supuesto de normalidad de v_2 (Wooldridge, 2010, pp. 587, 597).

Aún con estas advertencias claramente indicadas en los textos de econometría de referencia (v.g. Jeff Wooldridge), quienes se han dedicado a estudiar este tema en particular señalan que frente a los casos de modelos binarios con al menos un regresor endógeno binario siguen resolviéndose a través de los métodos de 2SLS, CF y LPM (Dong & Lewbel, 2015).

Esta es precisamente la situación en que tratan ambos problemas motivantes y la que discutiré en la segunda parte de este trabajo analizando opciones de diagnóstico y de solución ofrecidas por la bibliografía.

Determinantes de la pobreza multidimensional: modelo bajo sospecha.

En esta segunda parte de este trabajo presentaré el estudio sobre endogeneidad para el caso particular de un modelo de determinantes de la pobreza presentado por Fernández, Borrás & Ezquerro (Fernández, Borrás, & Ezquerro, 2017). Los autores tuvieron como objetivo contribuir a la agenda explicativa de la pobreza desde un punto muy particular: responder a la pregunta de si al cambiar la medida de pobreza –de unidimensional a multidimensional- cambiaría o se mantendría la estructura de determinantes de la misma. Para ello utilizaron como variables dependientes la pobreza de ingresos estimada por el Instituto Nacional de Estadística de Uruguay (INE, 2009) y la propuesta de medición multidimensional presentada en Borrás (Borrás, 2017). Este ejercicio tuvo varios problemas metodológicos que cuestionan la especificación adoptada, varios de los cuales fueron presentados y revisados en otra parte (Fernández & Borrás, 2017).

Para ello, y en función de la revisión de trabajos antecedentes, se propone un modelo explicativo que incluye determinantes de la pobreza en dos niveles de análisis: el hogar –determinantes asociados a características sociales, económicas y demográficas de los hogares- y el individual –determinantes asociados a características de los integrantes del hogar. El esquema nº2 informa de los dos niveles, las dimensiones, la operacionalización y las hipótesis sobre endogeneidad planteadas. Obsérvese de paso, que hay 8 regresores sobre los que recaen sospechas de endogeneidad, cinco relativas al capital humano y tres a titularidades sociales.

Esquema nº2. Determinantes de la pobreza especificados en el modelo completo, signo esperado e hipótesis sobre endogeneidad.

Dimensión	VARIABLES	Hipótesis	Endog?
Nivel de análisis hogar			
Titularidades de capital humano	En el hogar hay al menos un titulado de la Educación Superior (Universitaria, Formación Docentes y Terciaria No Universitaria)	Aversión	Si
Titularidades de capital económico	Propiedad del terreno y la vivienda que habita el hogar	Aversión	No
	El hogar dispone de otras propiedades inmuebles además de la vivienda	Aversión	No
	El hogar dispone de activos financieros	Aversión	No
	El hogar dispone de derechos de autoría, patentes, y similares	Aversión	No
Titularidad capital social	Algún miembro asiste o asistió a educación privada	Aversión	No
	Donaciones en dinero o especies de otros hogares o personas en el país	Aversión	No
	Remesas del extranjero	Aversión	No
	Al menos un miembro del hogar trabaja en una empresa de menos de 10	Riesgo	No
Titularidades de bienestar y protección social	En el hogar se percibe al menos una jubilación	Aversión	Si
	En el hogar al menos una persona que tiene empleo formal	Aversión	Si
	En el hogar al menos una persona tiene un empleo público	Aversión	Si
Posición estructural	Clases de ocupación según manual / no manual y no calificado / calificado. Referencia: Calificado-NoManual	Aversión	No
	Hogar con ocupaciones agrodependientes	Riesgo	No
Vulnerabilidad del hogar	Tipología de hogar según ciclo de vida del hogar. Referencia: hogar sin menores	Riesgo	No
	Perceptor afrodescendiente en el hogar	Riesgo	No
	Perceptor mujer en el hogar	Riesgo	No
Nivel de análisis individual			No
Titularidades de capital humano	Tiene título universitario	Aversión	Si
	Tiene título docente	Aversión	Si
	Tiene título terciario no universitario	Aversión	Si
	Tiene título técnico profesional (medio superior)	Aversión	Si
Titularidad de protección social	Atención de salud por FONASA	Aversión	No
Vulnerabilidad	Edad	Aversión	No
	Ascendencia racial afro. Categoría de referencia: no informa tener ascendencia racial afro alguna.	Riesgo	No
	Es mujer	Riesgo	No

Fuente: elaboración propia.



**XXXI CONGRESO ALAS
URUGUAY 2017**

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

El procedimiento seguido para la estimación de los modelos fue el siguiente. En primer lugar, se ajustaron modelos de regresión logística binaria para cada una de las variables que integran los niveles de análisis de hogar e individuo, seguido de ello, siguiendo el método de stepwise regression, se agrega bloque de variables por paso. A continuación se presentan los resultados del modelo general, computados como Average Marginal Effects (AVE) el cual consta con seis grupos de variables de nivel hogar y tres de nivel individual, tal como se especifica en la tabla n°1. Uno de los resultados más notorios es que la significación de las variables y el signo son similares en la pobreza multidimensional y en la de ingresos, salvo excepciones.

Tabla n°1. Efectos parciales promedio (AVE) de los modelos logísticos binarios ajustados.

Dimensión	Variable	LP Efecto parcial promedio	Sign.	PM Efecto parcial promedio	Sign.
Capital humano hogar	Hogar con al menos 1 miembro titulado de Educación Superior	-0.067	***	-0.047	***
Capital económico del hogar	Hogar propietario vivienda	-0.044	***	-0.027	***
	Hogar con activos financieros	-0.062	**	-0.014	NS
	Hogar con otras propiedades	-0.078	***	-0.020	***
Capital Social del hogar	Un miembro asistió Ed.Priv.	-0.042	***	-0.044	***
	Hogar recibe donaciones	-0.027	***	0.000	NS
	Hogar recibe remesas	0.026		-0.037	NS
Protección social hogar	Un miembro trabaja Peq./ Med empresa	0.001		0.031	***
	Al menos un empleo formal	-0.070	***	-0.021	***
	Al menos un empleo público	-0.076	***	-0.024	***
Posición en la estructura económica del hogar	Al menos una jubilación	-0.055	***	-0.020	***
	Ref. ISCO 1, 2 y 3				
	ISCO 4 y 5	0.029	***	0.016	***
	ISCO 7 y 8	0.017	***	0.022	***
Vulnerabilidad sociodemográfica del hogar	ISCO 9	0.075	***	0.029	***
	Al menos un perceptor en el sector agropecuario	-0.032	***	0.021	***
	Tipo de hogar (Referencia: sin menores)				
	Menor 6años /Adultos	0.081	***	0.037	***
	6a14años /Adultos	0.062	***	0.012	***
	De 0 a 14 años /Adultos	0.126	***	0.033	***
	14 18 y Adultos	0.055	***	0.013	***
	Menores y Adultos	0.185	***	0.046	***
	Menores, abuelo y adultos	0.104	***	0.018	***
	Todas las edades	0.110	***	0.03	***
Titularidad Capital humano individuo	Al menos un perceptor mujer	-0.035	***	0.01	***
	Al menos un perceptor afro	0.011	***	0.01	***
	Título Educación Superior ((Categoría de referencia Sin título)				
	Terciario No universitario	-0.034	NS	-0.01	NS
Protec. Soc.	Formación Docente	0.022	NS	-0.01	NS
	Universitario	0.006	NS	-0.03	***
	Título Técnico Medio Superior	-0.043	***	-0.03	***
Vulnerabilidad sociodemográfica individual	Tiene FONASA	-0.068	***	-0.07	***
	Afrodescendencia (Categoría de referencia no afro)			0.00	
	Afro no principal	0.016	***	0.00	NS
	Afro Principal	0.037	***	0.02	***
	Es mujer	0.002	NS	-0.02	***



XXXI CONGRESO ALAS URUGUAY 2017

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

	Edad (años)	-0.0004	***	0.001	***
--	-------------	---------	-----	-------	-----

Fuente: elaboración propia en base a ECH, 2015. (*) $P(\alpha) \leq 0.10$; (**) $P(\alpha) \leq 0.05$; $P(\alpha) \leq 0.01$: NS $P(\alpha) > 0.10$

Vistos estos resultados, y en la medida en que otorgan una respuesta muy importante a las interrogantes planteadas en el paper, corresponde discutir si la estimación es válida en presencia de endogeneidad. Esto es, si realmente están correctamente estimados los efectos de los determinantes. El mayor desafío se aclara, está en el problema del modelo de determinantes de la pobreza multidimensional. Es razonable suponer que tanto el tipo de endogeneidad resultante de un esquema de determinación simultánea como de autoregresión están presentes en el modelo explicativo original. Detallaré esta afirmación para luego ingresar en los aspectos empíricos y concluir con algunas reflexiones más generales.

Endogeneidad teórica de los regresores

La hipótesis sobre endogeneidad del capital humano es muy razonable y para nada sorprendente. Existe un extendido consenso en diferentes teorías sociológicas respecto de que el Capital Humano detentado por un individuo, y el conjunto de titularidades sociales utilizadas como regresores de la condición de pobreza, está asociada a la posición de clase social. Análisis elementales de correlación bivariada entre estos regresores permiten descartar la hipótesis nula, haciendo verosímil la representación de una situación de omisión de variable, si es que la clase social no fuera incluida en la pobreza⁶ (figura A en el esquema nº3). Podría decirse además que estos son resultados actuales de la acumulación educativa lograda en un período de tiempo prolongado. A su vez, tanto la posición actual de clase, como el capital humano, económico y social que detenta un individuo son un resultado o función de la posición de clase que el hogar de origen tenía, digamos por ejemplo, durante la primera infancia del individuo; es un caso de autoregresión (figura B en el esquema nº3). Vistos estos elementos, podría decirse que no resulta ignorable; un problema distinto es si puede hacerse una prueba concluyente o aún, si pueden haber soluciones económicas robustas.

⁶ No son despreciables las correlaciones de Pearson entre los activos financieros del hogar, la asistencia (actual o pasada) al sector privado de algún miembro de la familia, la tenencia de títulos educativos en algún miembro o y la titulación universitaria.



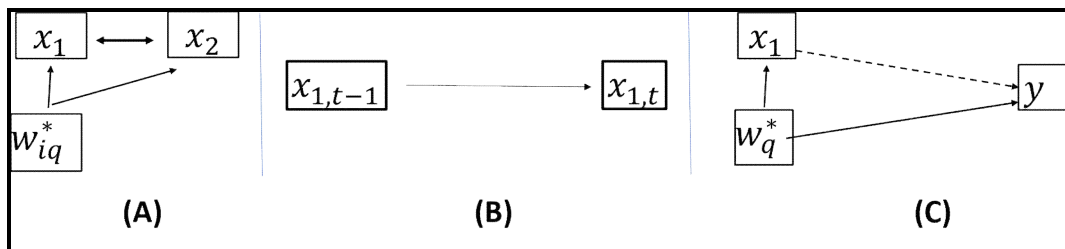
XXXI CONGRESO ALAS URUGUAY 2017

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

Esquema nº3. Tres formas de endogeneidad.



Sin embargo, también existe otra causa de endogeneidad, muy propia y específica de uno de los motivadores de los estudios sobre determinantes: la inclusión de las dimensiones de logro educativo y de protección social (empleo formal y jubilación) como dimensión constitutiva de la medida multidimensional de la pobreza. Las titularidades de capital humano detentadas por un individuo son una función de los años de educación formal aprobados. A su vez, la pobreza multidimensional está identificada toda vez que los años de educación formal aprobados son menores al umbral de 6 o 9 años, dependiendo de la edad de la persona. Dado que hemos construido la matriz de información sobre los determinantes de la pobreza, X siguiendo el procedimiento de atribuir las propiedades del globales, estructurales y analíticas del hogar a todos sus individuos, tendemos por esta vía una fuente de endogeneidad en la explicación de la pobreza del tipo de la presentada en la figura C del esquema nº2. La omisión de la escolaridad, relacionada tanto con la matriz X como con la matriz Y , generarían endogeneidad. Es cierto que la relación no es directa: los puntos de corte en la escolaridad aprobada son distintos en uno y otro caso y además, la medida de Pobreza Mutinomial adoptada aquí (Fernández, Borrás, & Ezquerra, 2017), al obviar el criterio de la unión y adoptar un valor de $C=0.40$, impide que una persona sea pobre por única razón de la dimensión educativa.

Una tercera situación en que configura endogeneidad es más compleja. Dos son las variables que intervienen. Es pobre multidimensional si algún miembro del hogar está ocupado por más de tres meses y no tiene cobertura de seguridad social. Es pobre multidimensional una persona si en su hogar hay un miembro mayor de 65 años que no perciba una jubilación, estando a su vez inactivo. A su vez, uno de los determinantes más sustantivos que hemos relevado de la bibliografía es la protección de seguridad social en el trabajo (independiente del tiempo transcurrido). A no ser por el requisito del tiempo (los tres meses), la función sería la misma que está incorporada a la matriz Y y a la matriz X .

En síntesis, el problema de marras trata con un modelo de determinantes original que cuenta con al menos 8 regresores endógenos, ambos de carácter binario. Esto ubica al problema dentro de la situación más compleja y controversial de tratamientos de la endogeneidad considerados por la bibliografía; el caso tipo IV.

Primer ejercicio: prueba estadística aplicada al modelo de determinantes



**XXXI CONGRESO ALAS
URUGUAY 2017**
3 - 8 Diciembre / Montevideo
Las encrucijadas abiertas de América Latina
La sociología en tiempos de cambio

La primera forma simple y recomendada para analizar empíricamente este tema es a través del examen de la matriz de correlaciones de Pearson. Las tablas 2 y 3 presentan este análisis.

Tabla 2. Correlaciones r de Pearson observadas entre los años de escolaridad y las variables incluidas en las matriz de explicación y la matriz de identificación multidimensional de la pobreza

	Adescol
Años de escolaridad (adescol) (variable omitida, individuo)	1.000
En el hogar hay al menos un titulado de la Educación Superior (njdegree)	0.410
Matriz X: propiedades absolutas	
Título Técnico Medio Superior (edutec2rr)	0.175
Título Terciario No Universitario (aces2_2)	0.146
Título de Formación Docente (maestro Primaria o profesor Media) (aces2_3)	0.256
Título de Educación Universitaria (aces2_4)	0.466
Matriz Y: propiedades absolutas	
Adulto que no aprobó el ciclo de Educación requerido por Ley en su momento	-0.286
Menor que no asiste debiendo asistir según su edad conforme a la Ley	-0.053
En el hogar hay al menos un adulto que no aprobó el ciclo de Educación exigido por Ley en su momento	-0.419

Fuente: elaboración propia con microdatos INE de la ECH 2015, base fusionada.

Las magnitudes son entre bajas y moderadas para las celdas que más atención deben merecer en ambas tablas. En negritas se destacan aquellas más preocupantes. Es claro que existe multicolinealidad entre las variables individuales y agregadas especificadas como determinantes de la pobreza: así por ejemplo entre la titulación universitaria o de formación docente con los años de escolaridad y con el número de títulos de educación superior (tabla nº2). Dada la magnitud de estas correlaciones, parecería no razonable introducirlas o buscar otra especificación que redujera el problema. La situación de endogeneidad mostrada por la columna “adescol” (variable omitida, candidata a instrumento) sería de mayor entidad con la titulación universitaria o la titulación de formación docente, y el número de títulos terciarios y la condición de privación educativa entre los adultos del hogar. Esta evidencia hace plausible la hipótesis de que coexistirían dos tipos de endogeneidad aquí, siendo la *omisión* de variable por construcción conceptual el más importante a tratar.

La situación en el caso de las variables de seguridad social es algo distinta, empezando porque la magnitud del problema aparentaría ser menor. La relación entre las dos matrices de información, Y y X , está reportada en el primer panel de la tabla 3. Se puede observar que la identificación de un hogar como pobre (y por tanto, de todos sus miembros) depende fuertemente de que en el hogar exista un adulto ocupado con más de tres meses de antigüedad en su empleo principal declarado y que no hace aportes a la seguridad social (BPS): $r=0.861$. Pero, esta condición tiene solo una relación baja con la variable analítica de la



matriz **X** (hforemp): $r=-0.220$, así como con la propiedad absoluta de ser empleado público (pubemp): $r=0.107$. La relación entre la condición de adulto con empleo formal (foremp) y la variable analítica del hogar (hforemp) es más débil que la observada al identificar pobreza (matriz **Y**) aunque es de magnitud moderada: $r=0.466$. A su vez, esta propiedad analítica tiene una relación baja con la condición de empleado público: $r=0.180$. Finalmente, cabe señalar que la existencia de una correlación moderada entre la condición de empleado formal y la condición de empleado público, $r=0.385$. Parecería razonable sostener que la variable analítica del hogar que informa de la tenencia de una titularidad de empleo formal, sería *endógena por variable omitida* a la tenencia de un empleo formal o un empleo público en alguno de sus integrantes. Esta última variable parecería ser una candidata interesante a instrumento.

Tabla 3. Correlaciones r de Pearson observadas entre los años de escolaridad y las variables incluidas en las matriz de explicación y la matriz de identificación multidimensional de la pobreza

	Hforemp	tss_h	foremp
[1] Propiedades analíticas atribuidas a los miembros como propiedades contextuales			
Adulto ocupado con más de tres meses de antigüedad en su empleo principal sin aportes a la seguridad social	-0.220	0.861	-0.230
Al menos un miembro del hogar sufre de una privación de seguridad social (tss_h)	-0.220	1	-0.271
Al menos un miembro del hogar tiene un empleo formal (hforemp)	1	-0.220	0.466
Al menos un miembro del hogar tiene una jubilación	-0.331	-0.020	-0.193
Al menos un miembro del hogar tiene un empleo público	0.312	-0.098	0.201
[3] Matriz X: propiedades absolutas			
Tiene ingresos por jubilación (nacional o extranjera) (jubil)	-0.382	-0.032	-0.260
Tiene un empleo público (pubemp)	0.180	-0.107	0.385

Fuente: elaboración propia con microdatos INE de la ECH 2015, base fusionada.

Estrictamente, las pruebas mediante correlaciones de Pearsons son elementalmente parciales por dos razones fundamentales. La primera es la naturaleza lineal de la relación capturada por el coeficiente de Pearsons, que podría ser el caso, subestimara la magnitud de la verdadera relación entre las variables, conduciendo a una conclusión equivocada sobre la incidencia que pudiera tener el problema. La segunda limitación tiene que ver con la propia especificación del problema establecido en la ecuación [11]: el testeo sobre la omisión de variables (y por ende de la calidad de los instrumentos) está formulado como efecto marginal por lo que debería ajustarse una correlación parcial más que una correlación bivariada.

Segundo ejercicio: estimación por CF



**XXXI CONGRESO ALAS
URUGUAY 2017**
3 - 8 Diciembre / Montevideo
Las encrucijadas abiertas de América Latina
La sociología en tiempos de cambio

La siguiente estrategia consiste en ajustar el modelo reducido por CF en ambos casos, esto es, con un regresor binario, pero en lugar de utilizar el comando ivprobit del STATA⁷, re-estimaré el modelo original (con endogeneidad) mediante LPM en ambas ecuaciones (estructurales y reducidas).

Este método se aplicó en 6 oportunidades, tres para las variables de titularidades de seguridad social y tres para variables de credenciales de capital humano, variables dicotómicas sospechosas de endogeneidad en el modelo explicativo de la pobreza multidimensional (véase esquema nº6.3). Para cada caso, para el primer caso, se usó como instrumento la tenencia de un empleo público, en tanto para los segundos casos, se usó el número de años de escolaridad formal aprobados. Las seis ecuaciones reducidas se ajustaron por LPM, asumiendo la limitante indicada arriba y se computaron los residuos estandarizados. Luego la ecuación [12] se especificó únicamente con el residuo de cada variable endógena examinada. En este sentido la prueba es parcial y no permite concluir sobre otras potenciales endogeneidades.

Tabla 4. Prueba de exogeneidad de regresores por el método de función de control (CF). Determinantes pobreza multidimensional (IV LPM)

	Hogar con al menos 1 empleado público	Hogar con al menos 1 empleado formal	Hogar con al menos 1 jubilación	Título universitario	Título FD	Número de títulos terciarios
[1] R2 en la ecuación reducida	0.445	0.407	0.382	0.435	0.246	0.597
[2] IV (empleo público)	0.772***	0.181***	- 0.108 ***			
[3] IV (adescol)				0.0204 ***	0.010 ***	- 0.002 ***
[4] Coeficiente (sesgado) var endógena	-0.646***	-0.557***	-0.646 ***	-1.120 ***	-0.147	- 1.337 ***
[5] Coeficiente var. endógena en ecuación estructural	-0.003*	-0.056***	-0.035 ***	-0.001	0.010 ***	-0.001
[6] coeficiente residuo en ecuación estructural	-0.027***	-0.705***	0.191 ***	-0.001	-0.003	-0.003
[7] DWH test (χ^2)	77.05 ***	189.45***	189.45 ***	0.03	0.03	0.034

Los resultados, presentados en la tabla 4 informan que la endogeneidad estaría presente con especial entidad en las variables relativas a las titularidades de seguridad social y a la titulación universitaria. Luego se ensayó la estimación de seis ecuaciones estructurales conforme [11], una por variable endógena, incorporando los residuos computados en la primera etapa a la ecuación estructural. El propósito fue testear si $\rho_1 = 0$ mediante el Test de DWH (línea 7).

Los hallazgos son diversos. Las titularidades sociales tienen endogeneidad. En las tres se descarta la hipótesis nula con estadísticos DWH altamente significativos. Sin embargo, y tal como se aprecia, no habría

⁷ Despite the name, ivprobit is not an iv estimator. It's a control function. (Baum, Dong, Lewbel, & Yang, 2012, pág. 818)



XXXI CONGRESO ALAS URUGUAY 2017

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

modificación de la significación y el sentido de la asociación de la variable endógena examinada en cada modelo. Es decir, a juzgar por estos *resultados parciales*, los estimados estarían sesgados solo en su magnitud.

Una conclusión diferente cabría observar en las variables de capital humano especificadas a nivel del hogar. Observando el Test DWH, se puede descartar la endogeneidad en cuanto a las tres variables. Sin embargo, en la ecuación estructural el resultado informa de un cambio en la significación además de un cambio también en el sentido de la relación. Estos resultados no parecerían ser razonables. Esto obliga a volver sobre la bibliografía. Precisamente, ésta advierte que, más allá de las ventajas que le son reconocidas a este método, *“CF generally cannot be implemented or is inconsistent when any endogenous regressor in X is discrete, censored, limited, or otherwise not continuously distributed”* (Lewbel, 2011). Más contundente aún *“Control function estimators are typically consistent only when the endogenous regressors are continuously distributed (because one cannot otherwise estimate the latent error e), and so should not be used when the endogenous regressors are discrete or limited”* (Dong & Lewbel, 2015).

Discusión de resultados y conclusiones

En el artículo mi objetivo es aportar a la agenda de los estudios de la pobreza pero desde un interés metodológico centrado en su explicación. Estoy convencido que este objetivo es tan importante como la correcta identificación y medición del fenómeno, temas a los que se han hecho anteriormente contribuciones de gran relevancia y que han acaparado la atención académica. Sin embargo, y esta es la hipótesis que nos orienta en este y en otro trabajo, la explicación de la pobreza resulta fundamental en este contexto de aparente estancamiento en la agenda de las políticas sociales, toda vez que se ha logrado en la última década, disminuir la incidencia de esta deuda. Discutir modelos explicativos, por lo tanto, da insumos sustantivos para idear estrategias para superarla.

El análisis aquí presentado tiene que ver con la robustez de los hallazgos de una investigación publicada por el autor y otros dos colegas que, a juicio de los revisores, padece de varios problemas de endogeneidad. La importancia de este tema en la econometría ha ido creciendo desde mediados de la década pasada hasta convertirse en uno de los más importantes, junto con el problema del sesgo de selección, para la agenda de las inferencias causales con base en datos cuantitativos.

En el modelo de marras, di cuenta tanto de endogeneidad entre las medidas de pobreza (sobre todo la multidimensional, pero no solo) y los determinantes, así como de estos últimos entre sí. Se hizo un primer ejercicio exploratorio con base en correlaciones, cuestión que permitió mejorar la comprensión de los problemas. Luego se puso en práctica un método, aquel que la bibliografía sugiere de mayor utilización para este caso que es el ajuste de modelos lineales de probabilidad con variables instrumentales (IV LPM) los cuales resultan de la adaptación del método de la función de control (CF). Aquí se hizo en forma parcial, ajustando seis modelos de este tipo.



XXXI CONGRESO ALAS URUGUAY 2017

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

Las pruebas de hipótesis permiten descartar la hipótesis nula de la exogeneidad en tres de los seis regresores sospechosos y no lo permiten en otros tres. Los primeros se relacionan con medidas de la dimensión “titularidades sociales” y tienen por base tres aspectos: empleo formal, empleo público y jubilaciones. Los tres restantes casos remiten a credenciales educativas que hacen al capital cultural del hogar: título universitario, título docente y otro título terciario. Aún en el caso de la presencia de endogeneidad no ignorable, se observó que los coeficientes estimados por este método mantienen la significación y el mismo sentido que aquel estimado originalmente por un modelo logístico ignorando la endogeneidad. El cambio se halló en la magnitud, la cual fue reducida drásticamente. Los demás coeficientes estimados no sufrieron cambios.

El ejercicio realizado resulta observable económicamente por dos razones. La primera elemental es que viola el supuesto de la distribución requerida en el regresor para aplicar este método. Varios autores advierten sobre el uso de este método en este caso. Sin embargo, los métodos específicos tampoco encuentran gran aceptación en la bibliografía. La segunda observación es que por razones de ajuste, aquí se hicieron ejercicios de CF independientes, uno por variable sospechosa, cuando parecería más razonable hacer un único ejercicio. Sin embargo, hacer esto implica otros problemas económicos mayores: por un lado, contar con un mayor número de “buenos instrumentos”, cuestión que a su vez genera mayores restricciones a probar, ineficiencias por lo tanto, y entre otras consecuencias, incrementa la multicolinealidad entre los residuos incorporados en la ecuación de regresión principal.

Llegado a este punto de la investigación, conviene por lo tanto hacer una pausa, reportar los resultados tal como se ha hecho, y pasar al análisis metodológico sobre las medidas que se están examinando. Aquí, tal como dijo una de las colegas que ha estado involucrada en estos análisis, hay un problema derivado del modelo teórico, o mejor aún de su operacionalización. Su diagnóstico es básicamente correcto: la endogeneidad discutida en estos ejercicios se ha generado en la medida en que las variables regresoras sospechosas son producto de “recodificaciones” o de “combinaciones” de información proveniente de los individuos miembros de los hogares. Información que además, también es utilizada para construir una de las medidas de la pobreza, la multidimensional.

Retomaré esto brevemente que fuera planteado en el esquema nº3, figura C. Los años de escolaridad aprobados por un individuo están a la base de la generación de todos los indicadores de credencial educativa y de satisfacción del estándar de educación obligatoria; es decir, todos los indicadores excepto uno (la titulación de la educación media técnica) están originados en esta variable. Lo interesante es que en estos casos, las pruebas no permiten descartar la hipótesis nula, con lo cual cabría concluir que la especificación adoptada cubre razonablemente bien la varianza en la variable original. La situación con las tres variables de titularidad social es lógicamente semejante, pero el resultado empírico es diferente: los indicadores de nivel individual excluidos de la ecuación principal generan esperanzas en las distribuciones condicionales del error distintas de cero. La información excluida es relevante. Es decir la información *individual* aún tiene un peso que no es correctamente representada por la *información colectiva* en una medida en que esta información ingresa también como un elemento constitutivo de la variable *dependiente*. Ele-



XXXI CONGRESO ALAS URUGUAY 2017

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

mento que además tiene un peso superlativo importante, tal como Borrás lo ha indicado en su trabajo (Borrás, 2017).

El punto llegado aquí es a mi juicio conceptual y tiene que ver con las respuestas que se le otorgue a la siguiente pregunta: ¿la existencia de la compleja endogeneidad identificada en el modelo invalida la agenda de la explicación teórica de los determinantes de la pobreza multidimensional al punto que haya que abandonar este camino explicativo?

Referencias

Baum, C., Dong, Y., Lewbel, A., & Yang, T. (2012). Binary Choice Models with Endogenous Regressors. *Stata Conference 2012*. San Diego: Stata Incorporated.

Borrás, V. (2017). Multidimensionalidad de la pobreza en Uruguay, 2006 - 2013. Análisis de Cuatro Dominios Geográficos. *Revista de Ciencias Sociales*, 30(40), 13-42.

Cameron, C., & Trivedi, P. (2010). *Microeconometrics using Stata. Revised Edition*. College Station, TX: The Stata Press.

Dong, Y., & Lewbel, A. (2012). *Simple Estimators for Binary Choice Models with Endogenous Regressors*. Irvine, CA: University of California at Irvine, School of Social Sciences, Economics.

Dong, Y., & Lewbel, A. (2015). A simple estimator for binary choice models with endogenous regressors. *Journal of Econometric Theory*, 34(1-2), 85-105.

Fernández, T., & Borrás, V. (2017). Notas para el estudio de los determinantes de la pobreza. En AAVV, *El Uruguay desde la Sociología. Volumen 16*. Montevideo: Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de la República.

Fernández, T., Borrás, V., & Esquerria, P. (2017). *Determinantes de la pobreza: modelo teórico y estimación para indicadores monetarios y no monetarios. Uruguay 2015*. Montevideo: Documento de trabajo. Grupo de Investigación sobre Desigualdad, Territorios y Transiciones. Departamento de Sociología, Universidad de la República.

Fernández, T., Borrás, V., & Esquerria, P. (2017). Pobreza de ingresos y pobreza multidimensional. Una comparación de sus determinantes para Uruguay 2015. *Sociedad*, 37, 155-186.

Fernández, T., Marques, A., & Lorenzo, V. (2017). Educación Media Técnica y la inserción laboral entre los 16 y los 25 años en Uruguay. Un estudio con base a los microdatos del Panel PISA 2003-2012. Montevideo: Discussion Paper. Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de la República.

Greene, W. (2012). *Econometric Analysis. Seventh Edition*. New York: Prentice Hall.

Gujarati, D. (2004). *Econometría. Cuarta Edición*. México, D.F.: McGraw-Hill.

INE. (2009). Línea de pobreza e indigencia 2006. Metodología y resultados. Uruguay.



**XXXI CONGRESO ALAS
URUGUAY 2017**

3 - 8 Diciembre / Montevideo

Las encrucijadas abiertas de América Latina

La sociología en tiempos de cambio

Lokshin, M., & Sajaia, Z. (2011). Impact of interventions on discrete outcomes: maximum likelihood estimation of binary choice models with endogenous regressors. *Stata Journal*, 11(3), 368-385.

Wooldridge, J. (2010). *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: The MIT Press.