

Uso de paneles sintéticos para estimar movilidad intergeneracional*

James E. Foster** Jonathan Rothbaum***

Resumen

Utilizando distintas medidas de movilidad, tales como curvas de movilidad, IGE, correlación y matrices de transición, se muestra que los paneles sintéticos pueden utilizarse para estimar con razonable precisión la movilidad intergeneracional de ingreso ante la ausencia de datos de panel. Al hacer esto, se introduce el uso de cópulas para mejorar la precisión de las estimaciones de movilidad por el método de panel sintético sobre los métodos que imponen fuertes supuestos de distribución y cuya invalidez se reconoce ampliamente. Se valida el uso de paneles sintéticos basados en cópulas sobre la movilidad intergeneracional e intrageneracional, y se estimar la movilidad de ingreso en el contexto de México, donde no existen datos de panel. Fue posible construir así estimaciones para cada una de las medidas de movilidad de ingreso en México, cuantificando las brechas en la igualdad de oportunidades, por ejemplo al comparar los prospectos de hijos nacidos en los hogares más ricos con aquéllos nacidos en los hogares más pobres: se encontró que los hijos de los hogares del decil más pobre tienen una probabilidad 11 veces mayor de quedarse ahí en su vida adulta que los que crecen en el decil más alto, de convertirse en adultos en el decil más bajo. También resulta 10 veces más probable que los hijos que provienen de los hogares del decil más rico se posicionen en su vida adulta en el decil más alto que los que nacieron en el decil más pobre. Finalmente, los resultados sugieren que la movilidad por rangos es muy similar en México y en Estados Unidos.

Palabras clave: Desigualdad, Movilidad Intergeneracional, pseudo-panel data, México.

Códigos JEL : D63, J62, C23

* Este artículo fue publicado bajo el mismo título como Capítulo XII en: Vélez Grajales R., J.E. Huerta Wong y R. M Campos Vázquez (Eds).2015. *México ¿ El motor inmóvil?* (pp 501-621). Se agradece la autorización de los autores y del Centro de Estudios Espinosa Yglesias para que el artículo pueda ser reproducido en *SobreMéxico. Temas de Economía* y así coadyuvar a la diseminación de metodologías útiles para la investigación aplicada sobre temas de movilidad social.

** Departamento de Economía e Instituto de Política Económica Internacional, George Washington University.

*** Instituto de Política Económica Internacional, George Washington University.

1. Introducción

Desigualdad y desigualdad de oportunidades son temas de política pública de suma importancia en muchos países alrededor del mundo. Ha habido una gran cantidad de investigación académica que documenta los altos y profundos niveles de desigualdad en muchos países desarrollados y en desarrollo. Las políticas que apuntan hacia resultados educativos y de salud con frecuencia tienen como objetivo disminuir la desigualdad de oportunidades y nivelar la cancha de juego entre los hijos de los pobres y los ricos. Sin embargo, debido a los altos requerimientos de datos, sabemos muy poco, tanto del grado de igualdad de oportunidades y movilidad en la mayoría de los países, como de la forma en la que están cambiando con el paso del tiempo. Lo anterior es cierto, en especial, si hablamos de medidas de movilidad de bienestar distintas de las de ingreso, tales como consumo y riqueza.

Ha habido intentos para estimar la movilidad en los ingresos de padres e hijos sin recurrir a datos de panel. Sin embargo, estos estudios generalmente estiman la movilidad de una *proxy* del ingreso, tales como nivel educativo, clase ocupacional, índices de estatus o alguna combinación de todas éstas, mismas que sólo capturan parcialmente la relación entre ingresos de padres e hijos.¹ Otros trabajos estiman la elasticidad intergenera-

cional del ingreso en países desarrollados, por ejemplo, por medio de modelos de variables instrumentales en dos etapas para dos muestras,² o con base en casos donde los padres e hijos adultos cohabitan.³

Un problema similar existe en la literatura sobre movilidad intrageneracional. En ese contexto, la carencia de paneles que sigan a individuos u hogares durante lapsos cortos ha limitado la capacidad de los investigadores para estudiar transiciones relativas a entradas y salidas de pobreza, así como la pobreza crónica y transitoria. Para superar las limitaciones en cuanto a disponibilidad de datos, los trabajos de Deaton; Banks, Blundell y Brugiavini; así como Antman y Mckenzie han utilizado datos de múltiples encuestas de corte transver-

phy and Generations Intergenerational Educational Mobility in Post-Reform India», *Policy Research Working Paper* 6055, Banco Mundial, 2012; M. Emran y F. Shilpi, «Intergenerational Occupational Mobility in Rural Economy», *The Journal of Human Resources*, vol. 46, núm. 2, 2011, pp. 427-458; F. Torche, «Unequal But Fluid: Social Mobility in Chile in Comparative Perspective», *American Sociological Review*, vol. 70, núm. 3, 2005, pp. 422-450; J. Behrman, *et al.*, «Intergenerational Mobility in Latin America», *Economía*, vol. 2, núm. 1, 2001, pp. 1-44; M. Emran y Y. Sun, «Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 1988-2002», *Working Paper*, 2011.

2 S. Guimaraes Ferreira y F. Veloso, «Intergenerational Mobility of Wages in Brazil», *Brazilian Review of Econometrics*, vol. 26, núm. 2, 2006, pp. 181-211; C. Dunn, «The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil», *The BE Journal of Economic Analysis & Policy*, vol. 7, núm. 2, 2007, pp. 1-42; J. Núñez y L. Miranda, «Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile», *The BE Journal of Economic Analysis & Policy*, vol. 10, núm. 1, 2010, pp. 1-15.

3 T. Hertz, «Education, Inequality and Economic Mobility in South Africa», tesis doctoral, University of Massachusetts, 2001; D. Quheng, *et al.*, «Intergenerational Income Persistence in Urban China», *IZA Discussion Paper* 6907, Institute for the Study of Labor (IZA), 2012; V. Hnatkowska, *et al.*, «Breaking the Caste Barrier Intergenerational Mobility in India», *Journal of Human Resources*, vol. 48, núm. 2, 2013, pp. 435-473.

1 T. Hertz, *et al.*, «The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends», *The BE Journal of Economic Analysis & Policy*, vol. 7, núm. 2, 2007, pp. 1-48; A. Nimubona, y D. Vencatachellum, «Intergenerational Education Mobility of Black and White South Africans», *Journal of Population Economics*, vol. 20, núm. 1, 2007, pp. 149-182; M. Emran y F. Shilpi, «Gender, Geogra-

sal y cambios en los niveles de ingreso por cohortes para estudiar pobreza y movilidad.⁴

Con base en el trabajo de Elbers, Lanjouw y Lanjouw,⁵ Dang, Lanjouw, Luoto y McKenzie (a partir de aquí DLLM) propusieron un enfoque para estudiar transiciones de pobreza con base en paneles sintéticos que se conforman a partir de múltiples encuestas de corte transversal. En su método, el ingreso del hogar se estima usando características que no varían con el tiempo para datos de dos cortes transversales por separado. Para cualquier hogar observado en el corte transversal inicial —pero no en el final— la probabilidad de transitar fuera de la pobreza depende de su ingreso estimado en el periodo final, así como de la correlación entre el valor residual del ingreso estimado en los periodos inicial y final. Al hacer supuestos conservadores acerca de la relación entre los residuos del ingreso inicial y final, los autores establecen límites superiores e inferiores de la proporción de la población que escapa o cae en pobreza para cualquier posible línea de pobreza.⁶ En artículos subsecuentes, Cruces, *et al.*,⁷ y Dang y Lanjouw⁸ hacen supuestos más restrictivos acerca de la relación entre el ingreso inicial y el ingreso final, de modo que estimar la movilidad se reduce a estimar un parámetro simple. En un ejercicio de validación con datos de panel, muestran como esto les permite obtener estimaciones más precisas de los movimientos de entrada y de salida de la condición de pobreza.

La curva de movilidad es una herramienta útil para validar los resultados de las estimaciones de movilidad con base en paneles sintéticos, ya que permite comparar la estimación de la movilidad verdadera (o transiciones de pobreza) en todas las posibles líneas de pobreza de manera simultánea.⁹

4 A. Deaton, «Panel Data from Time Series of Cross-Sections», *Journal of Econometrics*, vol. 30, 1985, pp. 109–126; J. Banks, *et al.*, «Risk Pooling, Precautionary Saving and Consumption Growth», *The Review of Economic Studies*, vol. 68, núm. 4, 2001, pp. 757–779; F. Antman y D. McKenzie, «Earnings Mobility and Measurement Error: A Pseudo-Panel Approach», *Economic Development and Cultural Change*, vol. 56, núm. 1, 2007, pp. 125–161.

5 C. Elbers, *et al.*, «Micro-Level Estimation of Poverty and Inequality», *Econometrica*, vol. 71, núm. 1, 2003, pp. 355–364.

6 Hai-Anh Dang, *et al.*, «Using Repeated Cross-Sections to Explore Movements in and out of Poverty», *Policy Research Working Paper 5550*, Banco Mundial, 2011.

7 G. Cruces, *et al.*, «Intra-Generational Mobility and Repeated Cross-Sections: A Three-Country Validation Exercise», *Policy Research Working Paper 5916*, Banco Mundial, 2011.

8 Hai-Anh Dang y P. Lanjouw, «Measuring Poverty Dynamics with Synthetic Panels Based on Cross-Sections», *Policy Research Working Paper 6504*, Banco Mundial, 2013.

9 J. Foster y J. Rothbaum, «The Mobility Curve: Measuring the Impact of Mobility on Welfare», *Working Paper*, 2014.

En este capítulo, extendemos la investigación sobre paneles sintéticos de dos maneras. Primero, mostramos que una especificación semiparamétrica de cópula usando cópulas, aporta generalmente una estimación más precisa de la movilidad intra e intergeneracional que la especificación paramétrica en DLLM. Las especificaciones semiparamétricas de cópula las validamos usando una variedad de conjuntos de datos de los Estados Unidos. En particular, probamos estimaciones paramétricas y semiparamétricas para movilidad intrageneracional usando el «Panel Study of Income Dynamics Cross National Equivalence File» (PSID-CNEF), así como el archivo público del «Current Population Survey Annual Social and Economic Supplement» (CPS ASEC). Segundo, aplicamos la técnica de panel sintético para la estimación de la movilidad intergeneracional. Con las encuestas «National Longitudinal Surveys of Youth» (NLSY), mostramos que un modelo de cópula más flexible aporta una estimación más precisa de la movilidad intergeneracional conocida, en lugar de la especificación paramétrica propuesta por DLLM. Después de validar el uso de paneles sintéticos para estimar movilidad intergeneracional, aplicamos la técnica para medir la movilidad intergeneracional en México, donde no existen datos de panel que incluyan el ingreso emparejado de padres e hijos.

Para estimar la movilidad intergeneracional en México, utilizamos dos encuestas. La primera, la Encuesta ESRU de movilidad social en México (EMOVI),¹⁰ capturó información sobre las características demográficas y económicas de los hogares mexicanos en 2006 y 2011, además de formular una serie de preguntas retrospectivas acerca de las circunstancias de los entrevistados cuando vivían en los hogares de sus padres. Emparejamos un subgrupo de los hogares de la EMOVI con los hogares de la cohorte correspondiente a la de los padres de los entrevistados de ésta de la Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH).¹¹ Y así, construimos paneles sintéticos a partir de los datos transversales de la EMOVI y la ENIGH para estimar varias medidas de movilidad intergeneracional para mexicanos nacidos entre 1966 y 1981.

10 Encuesta financiada por la Fundación Espinosa Rugarcía (Fundación ESRU).

11 Encuesta a cargo del Instituto Nacional de Estadística y Geografía (INEGI).

2. Paneles sintéticos

La técnica de panel sintético propuesta por DLLM presenta varias ventajas sobre otras técnicas previas. Este método depende de menos supuestos que otros enfoques y permite hacer comparaciones al interior y entre grupos. Esto es relevante en el contexto de la movilidad intergeneracional y la igualdad de oportunidades en tanto que nos interese por las diferencias de movilidad por cuantil o raza, entre otras formas posibles descomposiciones.

Empezaremos con una breve discusión de la técnica de DLLM. Supongamos que hay dos encuestas aleatorias de N_1 y N_2 individuos respectivamente, que son cortes transversales aleatorios de las poblaciones de interés. Ésta podría ser la misma población de hogares para movilidad intrageneracional u hogares de padres e hijos para movilidad intergeneracional. Sea x_{it} un vector de características del hogar i en el periodo t . Estas características pueden no variar en el tiempo (raza, etnicidad, sexo, lugar de nacimiento, etc.), ser deterministas (edad) o sí variar en el tiempo, de tal manera que puedan recordarse con precisión (tipo de ocupación, estatus laboral, características del hogar tales como la propiedad de un automóvil, tamaño y características de las viviendas, etc.). Sea y_{it} una medida del estatus económico, tal como ingreso, riqueza o consumo, al cual para facilitar la exposición llamaremos ingreso. El ingreso puede expresarse como una proyección lineal de las características x_{it} tal como:

$$2.1 \quad y_{it} = \beta'_{it} x_{it} + \epsilon_{it}$$

Dada una línea de pobreza o un límite de ingreso c_t , la proporción de los hogares que experimentan movilidad ascendente puede definirse como:

$$2.2 \quad P(y_{i1} \leq c_1 \text{ y } y_{i2} > c_2)$$

Ésta es la proporción de la población por debajo de la línea de pobreza c_1 en el periodo 1, pero por encima de la línea de pobreza c_2 en el periodo 2. Para la movilidad descendente, $P(y_{i1} > c_1 \text{ y } y_{i2} \leq c_2)$ representa la transición de no pobre a la condición de pobre. Desafortunadamente, sin datos de panel no podemos observar y_{i1} y y_{i2} para los mismos hogares, de modo que no podemos calcular (2.2).

DLLM proponen reescribir la ecuación (2.2) y sustituir (2.1) en y_{it} de tal manera que:

$$2.3 \quad P(\epsilon_{i1} \leq c_1 - \beta'_1 x_{i1} \text{ y } \epsilon_{i2} > c_2 - \beta'_2 x_{i2}).$$

De esta ecuación, la movilidad ascendente depende sólo de la distribución conjunta de los términos de error, ya que todos los otros términos (c_t , β_t , x_{it} , ϵ_{it}) son conocidos o pueden ser estimados.

De (2.3), DLLM estiman los límites de la movilidad al hacer supuestos acerca de la relación entre ϵ_{i1} y ϵ_{i2} . Ellos estiman un límite superior de movilidad, al asumir que los términos de error son completamente independientes entre sí, con base en el método de *bootstrap* para estimar un nivel promedio de movilidad. Su estimación del límite inferior de movilidad se basa en la correlación perfecta de los términos de error que instrumentan asumiendo $\epsilon_{i1} = \epsilon_{i2}$ para cada hogar. DLLM confirman con datos de panel de Indonesia y Vietnam que la estimación verdadera está casi siempre contenida entre los límites, aun cuando la muestra se descomponga por subgrupos regionales. Sin embargo, los límites mismos pueden ser muy amplios. Por ejemplo, en la línea de pobreza en Indonesia entre 1997 y 2000, el nivel verdadero de movilidad ascendente en los datos de panel fue de 0.08, pero el límite inferior fue de 0.03 y el nivel superior fue de 0.12.¹² Fields y Viollaz usan datos de Chile para probar una variedad de conceptos de movilidad, incluyendo dependencia en el tiempo (correlación), movimiento de posiciones, movimiento compartido, movimiento de ingresos, y movilidad y desigualdad. Ellos también encuentran que los límites superiores e inferiores propuestos son demasiado amplios y aportan «información limitada acerca de las tasas de transición de la pobreza».¹³

Para obtener estimaciones más estrechas sobre los límites de movilidad o una estimación puntual, se requieren mayores supuestos acerca de la relación entre los residuales ϵ_{it} . Un supuesto posible, hecho por DLLM, es que $\epsilon_{i1} = \epsilon_{i2}$ siguen una distribución normal bivariada con un coeficiente de correlación ρ y desviaciones estándar $\sigma_{\epsilon 1}$ y $\sigma_{\epsilon 2}$. Uno puede obtener límites más estrechos asumiendo que hay una posible correlación máxima y mínima de los errores tal que $0 < \rho_L < \rho_H < 1$. Al reducir el rango de correlaciones posibles a $\rho_L = 0.3$ y $\rho_H = 0.7$, los límites sobre movilidad ascendente a través de las líneas de pobreza para Indonesia se reducen a [0.08, 0.12] en su artículo.¹⁴

12 Hai-Anh Dang, et al., op. cit.

13 G. Fields y M. Viollaz, «Can the Limitations of Panel Datasets Be Overcome by Using Pseudo-Panels to Estimate Income Mobility», *Working Paper*, 2013.

14 Hai-Anh Dang, et al., op. cit.

Cruces, *et al.* estiman la movilidad bajo el supuesto de normalidad bivariada usando datos de panel de Chile, Nicaragua y Perú.¹⁵ Ellos encuentran que «la metodología funciona bien en la predicción de la movilidad verdadera de entradas y salidas de la pobreza, a través de dos rondas de datos de corte transversal; la movilidad verdadera cae dentro de los dos límites la mayor parte del tiempo». Cruces, *et al.* también apuntan que mejorar el modelo (incluir más características en x_{it} para aportar una mejor estimación de y_{it} y por tanto residuales más pequeños), se traduce en estimaciones puntuales más precisas de movilidad a partir de paneles sintéticos, y límites más estrechos para un conjunto determinado de parámetros ρ_L y ρ_H .

Si el verdadero valor de ρ fuera conocido o pudiese estimarse, se podría obtener una estimación puntual para la movilidad en lugar de límites. Dang y Lanjouw proponen un método que usa cohortes de edad para estimar la correlación del ingreso entre el periodo 1 y el periodo 2 $\rho_{y_1y_2}$.¹⁶ La correlación del ingreso es igual a:

$$2.4 \quad \rho_{y_1y_2} = \frac{\text{cov}(y_{i1}, y_{i2})}{\sqrt{\text{var}(y_{i1})\text{var}(y_{i2})}} = \sqrt{\frac{\text{var}(y_{i1})}{\text{var}(y_{i2})}} \delta$$

Con base en el ingreso promedio para cada cohorte c en el periodo t , \bar{y}_{ct} , δ puede estimarse a partir de la regresión de cohortes:

$$2.5 \quad \bar{y}_{c2} = \delta \bar{y}_{c1} + \bar{v}_{c2}$$

Para que esta estimación no esté sesgada, la variable de cohorte debe satisfacer las condiciones de una variable instrumental (incluyendo exogeneidad y relevancia).

Con las correlaciones de ingreso estimado, la correlación entre los residuos de la regresión también puede estimarse mediante la reordenación de los términos en la ecuación de correlación usando (2.1) como:

$$2.6 \quad \rho_{y_1y_2} = \frac{\text{cov}(y_{i1}, y_{i2})}{\sqrt{\text{var}(y_{i1})\text{var}(y_{i2})}} = \frac{\text{cov}(\beta_1' x_{i1} + \epsilon_{i1}, \beta_2' x_{i2} + \epsilon_{i2})}{\sqrt{\text{var}(y_{i1}) \text{var}(y_{i2})}}$$

Si las dos muestras provienen de la misma población (con la misma distribución de x variables, un supuesto identificado en DLLM), entonces (2.6) puede reescribirse

como:

$$2.7 \quad \rho_{y_1y_2} = \frac{\beta_1' \text{var}(x_i) \beta_2 + \rho \sqrt{\sigma_{\epsilon 1}^2 \sigma_{\epsilon 2}^2}}{\sqrt{\text{var}(y_{i1}) \text{var}(y_{i2})}}$$

De (2.7), la correlación de los residuos de la regresión ρ puede estimarse dada una estimación de la correlación de ingresos $\rho_{y_1y_2}$ que como todos los otros términos puede ser estimada a partir de las distribuciones marginales de las variables.

Dang y Lanjouw prueban sus estimaciones de movilidad con las ρ 's estimadas con datos de un conjunto más amplio de países, incluyen a Bosnia-Herzegovina, Laos, los Estados Unidos, Perú, y Vietnam. Encuentran que sus resultados son «considerablemente precisos» y que «son buenos no solamente para la población en general, sino también para grupos más pequeños de población».¹⁷

Curvas de Movilidad

Con el fin de probar y validar diferentes técnicas de paneles sintéticos sobre la movilidad intergeneracional, primero utilizamos curvas de movilidad.¹⁸ Estas trazan las transiciones de entrada y salida de la condición de pobreza para todas las posibles líneas de pobreza. Esto nos permite comparar nuestros resultados con los de DLLM y analizar qué tan bien los paneles sintéticos predicen cambios en el ingreso en otros puntos en la distribución, y no solamente en la línea de pobreza o en un pequeño conjunto de líneas posibles.

En esta sección aportamos una breve síntesis de la curva de movilidad; misma que se define para comparar cómo las ganancias y las pérdidas en el ingreso afectan el bienestar. Con $y_t = (y_{it}, \dots, y_{nt})$ y $y = (y_1, y_2)$, la movilidad ascendente y descendente para una determinada cota c son:

$$3.1 \quad m_U(y, c) = \frac{1}{n} \sum_{i=1}^n I(y_{i1} \leq c) I(y_{i2} > c)$$

$$m_D(y, c) = \frac{1}{n} \sum_{i=1}^n I(y_{i1} > c) I(y_{i2} \leq c)$$

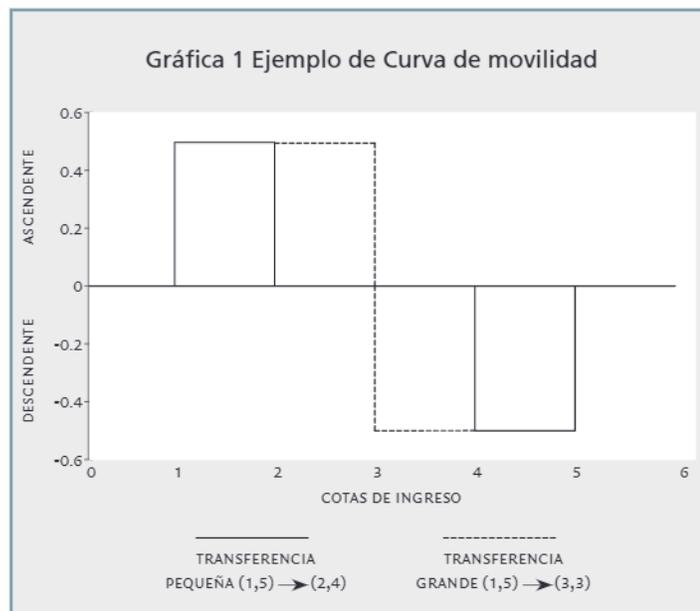
Bajo los supuestos de bienestar social utilitario y separable en el tiempo, la curva de movilidad se define de modo tal que si las ganancias en el ingreso resultaron en un mayor incremento per cápita en el bien-

15 G. Cruces, *et al.*, *op. cit.*

16 Hai-Anh Dang y P. Lanjouw, *op. cit.*

17 *Ibid.*, p. 32.

18 J. Foster y J. Rothbaum, *op. cit.*



Nota: Ejemplo de curva de movilidad con dominancia de primer orden en dos sociedades, A y B. Para cada cota (c) en el eje de la x , la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it+1} \geq c$) por encima del eje de la x , y la proporción de la población que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it+1} < c$) por debajo del eje de la x . La sociedad B tiene mayor movilidad ascendente que la sociedad A ($1 \rightarrow 3$ vs $1 \rightarrow 2$) lo cual también se muestra en la curva de movilidad por el hecho de que la movilidad ascendente para B es mayor o igual que la movilidad ascendente para A en todas las cotas. Lo mismo ocurre para la movilidad descendente. Por lo tanto, la movilidad de primer orden de B domina a A tanto en la movilidad ascendente como en la descendente, lo cual significa a su vez que las ganancias (pérdidas) en el bienestar se deben a que la movilidad ascendente (descendente) es mayor en B que en A.

Fuente: J. Foster y J. Rothbaum, «The Mobility Curve: Measuring the Impact of Mobility on Welfare», *Working Paper*, 2014.

estar de la sociedad B que en la sociedad A, entonces B experimentó más movilidad ascendente. Sea y^A y y^B los ingresos del periodo inicial y final para las sociedades A y B, respectivamente. De la ecuación (3.1), si $m_U(y^B, c) \geq m_U(y^A, c)$ para todas las posibles cotas c y para algunas cotas $m_U(y^B, c) > m_U(y^A, c)$, entonces B experimentó un mayor incremento per cápita en el bienestar de la movilidad ascendente para cualquier función de utilidad monótonamente creciente. Por definición entonces, la movilidad ascendente es más grande en B que en A. La curva de movilidad se construye al trazar m_U y m_D para todos los valores posibles de c . Al observar las curvas de movilidad para B y A, podemos fácilmente ver si B tiene más movilidad ascendente que A. La Gráfica 1 muestra un ejemplo tomado de Foster y Rothbaum con dos sociedades, A y B, que comparten ingresos idénticos en el periodo 1, $y_1^A = y_1^B = (1, 5)$, pero que difieren en el periodo 2, $y_2^A = (2, 4)$ y $y_2^B = (3, 3)$.¹⁹ Al revisar los números y las

curvas de movilidad en dicha gráfica, resulta claro que la sociedad experimentó mayor movilidad ascendente ($1 \rightarrow 3$ comparado a $1 \rightarrow 2$) y mayor movilidad descendente ($5 \rightarrow 3$ comparado a $5 \rightarrow 4$) que la sociedad A.

Como hemos señalado anteriormente, las curvas de movilidad son también una manera de observar las transiciones de la pobreza, ya que grafican las transiciones de salida de la condición de pobreza (movilidad ascendente) y de entrada a la condición de pobreza (movilidad descendente), a través de todas las posibles líneas de pobreza. Si $c_1 = c_2$, m_U en la ecuación (3.1) es igual a la ecuación (2.2).²⁰ Al permitirnos observar simultáneamente la movilidad a través de todas las posibles líneas de pobreza, las curvas de movilidad hacen más fácil validar la precisión de los paneles sintéticos en la estimación de la movilidad en todas las posibles líneas de pobreza, en lugar de una o un subconjunto de líneas escogidas arbitrariamente. Como tales, las curvas de movilidad aportan una forma de ver una «distribución» de la movilidad que no necesariamente es posible con un índice simple de medidas de movilidad o con matrices de transición.

3. Paneles sintéticos, errores normales bivariados, y curvas de movilidad

Dang y Lanjouw realizaron una serie de validaciones del uso de paneles sintéticos para estimar las transiciones de la pobreza.²¹ En esta sección discutiremos sus resultados. Dado que hemos extendido el uso de la técnica para la movilidad intergeneracional, también validamos la técnica de panel sintético para la movilidad intergeneracional utilizando las encuestas «National Longitudinal Surveys of Youth» (NLSY), para los periodos 1979 y 1997, de los Estados Unidos.

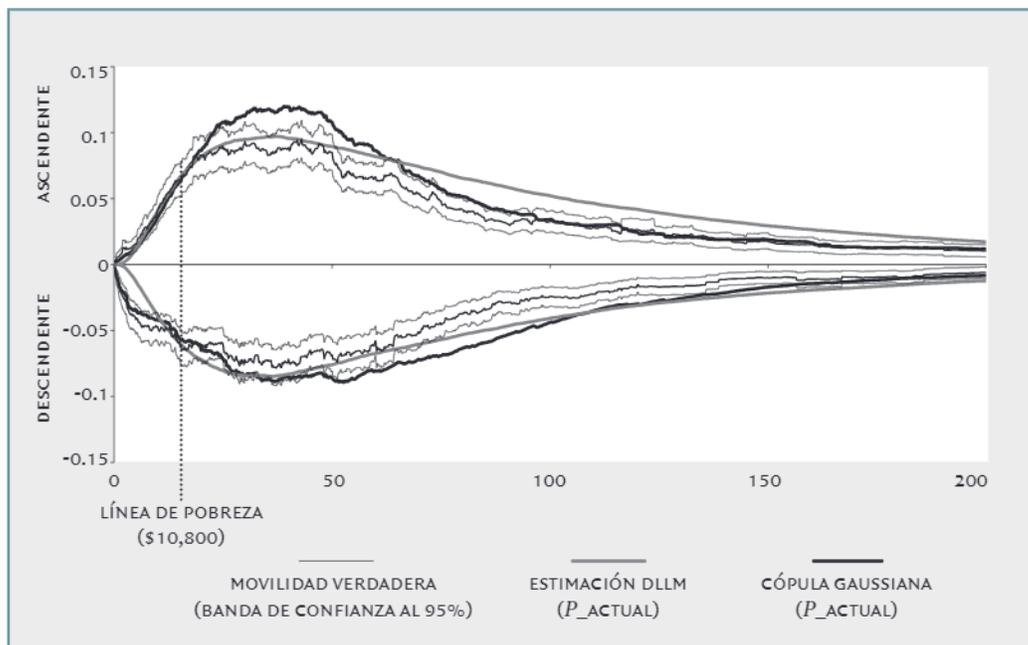
Dang y Lanjouw comparan las estimaciones del panel sintético de los movimientos de entrada y salida de la condición de pobreza en Bosnia-Herzegovina (2001-2004), Laos (2002/2003-2007/2008), Perú (2005-2006), los Estados Unidos (2007-2009), y Vietnam (2006-2008).²² Para 17 de las 20 transiciones comparadas (para cada caso: pobres

²⁰ Se requiere que las líneas de pobreza c_1 y c_2 no sean las mismas para cada hogar. Por ejemplo, si los ingresos del hogar son ajustados por su equivalencia, entonces la línea de pobreza para cada hogar sería la misma en dólares ajustados por equivalencia, pero no en ingreso absoluto.

²¹ Hai-Anh Dang y P. Lanjouw, *op. cit.*

²² *Ibid.* Los autores utilizaron la línea de pobreza del Panel Study of Income Dynamics (PSID) para los datos de los Estados Unidos. Para los datos de Bosnia-Herzegovina, usaron el percentil 20 del consumo para 2001. Los autores usaron la línea de pobreza oficial para los otros países.

Gráfica 2
Estimación sintética vs. Movilidad verdadera, PSID-CNEF 2005 y 2007



Notas: La curva de movilidad verdadera, ascendente y descendente, con intervalos de confianza al 95%, y las estimaciones DLLM y de panel sintético usando cópula Gaussiana, utilizan la correlación conocida de los errores MCO: ρ_{Actual} . Para cada cota (c) sobre el eje de la x , la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it} \geq c$) por encima del eje de la x , y la proporción que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it} < c$) por debajo del eje de la x . La estimación DLLM subestima la movilidad ascendente y descendente en las cotas de ingresos bajos y la sobrestima en las cotas de ingresos altos. Sin embargo, la estimación de las transiciones de entrada y salida de la pobreza (movilidad en la línea de pobreza) son muy precisas.
Fuente: Estimación propia con base en los datos del PSID-CNEF 2005 y 2007.

a pobres, pobres a no pobres, no pobres a pobres, no pobres a no pobres), las probabilidades de transición verdadera en los datos de panel y las estimaciones del panel sintético no fueron estadísticamente distintas entre sí.

Sin embargo, debemos tomar con cuidado estos resultados. A manera de ilustración, en la Gráfica 2 trazamos la curva de movilidad (muestral) verdadera junto con la estimación de panel sintético de la curva de movilidad de los Estados Unidos para los años 2004-2006, usando datos de los levantamientos de las rondas de 2005 y 2007 del PSID-CNEF. El panel sintético se estimó para individuos entre 25 y 55 años. Primero se calculó una regresión por MCO del ingreso con un número pequeño de variables independientes como edad, edad al cuadrado, años de educación, sexo y variables *dummy* para población de color e hispana.²³ Los resultados de la regresión por MCO aparecen en el Cuadro 1.

Para estimar ρ , usamos la medida de τ de Kendall (ρ_τ) para reducir el efecto de datos extremadamente atípicos. Bajo el supuesto de normalidad bivariada, la relación entre ρ_τ y ρ es: $\rho = \sin(\frac{\pi}{2} \rho_\tau)$.²⁴ Para el PSID-CNEF, la

correlación real entre los errores es de 0.65 y el estimado de ρ_τ es 0.78. Si no incluimos los valores atípicos cuyo ingreso predicho es 10 veces más alto o más bajo que su ingreso real, la correlación es 0.70.²⁵ Al estimar ρ de ρ_τ , tampoco tenemos que especificar qué valores atípicos remover debido a que ningún valor extremo simple tiene un efecto grande en ρ_τ .

En la línea de pobreza (\$10,790 en dólares de 2007), la predicción del panel sintético para la movilidad ascendente y descendente (que corresponde a las transiciones de pobre a no pobre, y de no pobre a pobre en el análisis de Dang y Lanjouw) es muy similar a la movilidad verdadera observada en el panel. Sin embargo, para cotas de ingresos medios y altos, la movilidad DLLM excede la movilidad verdadera con una banda de confianza de 95% obtenido por *bootstrapping* (a casi todas las cotas por encima de 50,000 dólares para movilidad ascendente y 70,000 dólares para movilidad descendente). Además, para niveles muy bajos de ingreso (por debajo de 4,000 dólares para movilidad ascendente y \$10,000 para movilidad descendente) DLLM subestima las transiciones de movilidad.

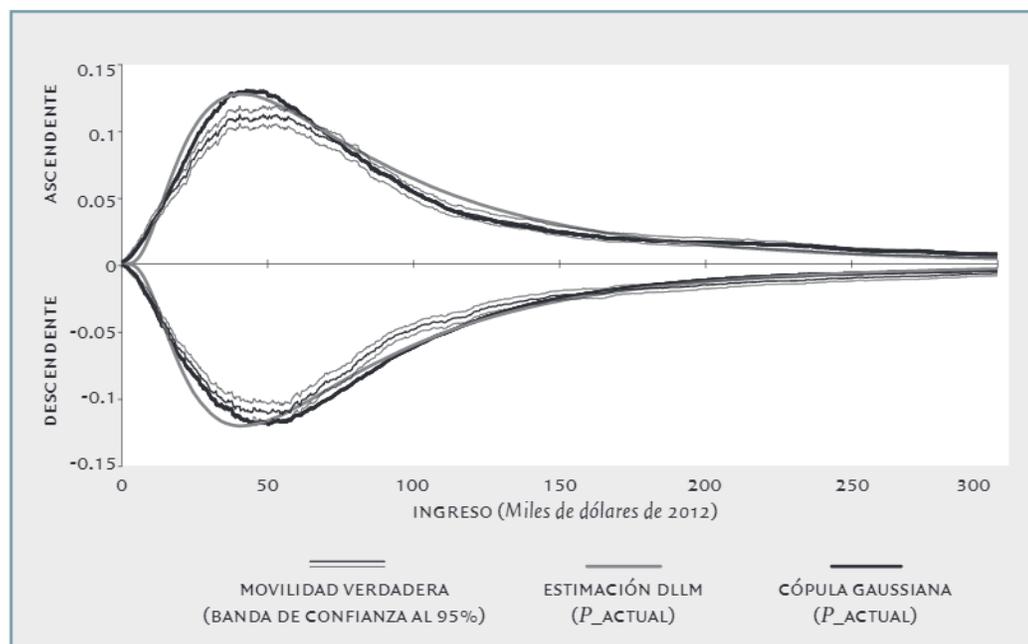
23 Estas variables son muy similares a las usadas por Hai-Anh Dang y P. Lanjouw, *op. cit.* con los datos del PSID. La meta no es replicar sus resultados con exactitud, sino mostrar cómo los paneles sintéticos pueden aportar una estimación sesgada de movilidad en muchos puntos de la curva de movilidad y también permitir resultados precisos de transiciones de pobreza.

24 S. Demarta y A. McNeil, «The T Copula and Related Copulas», *International Statistical Review*, vol. 73, núm. 1, 2007, pp. 111-129.

25 Como un ejemplo del caso de la movilidad intergeneracional, para la muestra NLSY-1997, al remover estos datos atípicos se incrementa la ρ de 0.200 a 0.227, y la correlación estimada de ρ_τ es 0.246.

Gráfica 3

Estimación sintética vs. Movilidad verdadera de ingreso, CPS ASEC (panel 2005-2006)



Notas: La curva de movilidad verdadera, ascendente y descendente, con intervalos de confianza al 95%, y las estimaciones DLLM y de panel sintético usando cópula Guassiana, utilizan la correlación conocida de los errores MCO: ρ_{Actual} . Para cada cota (c) sobre el eje de la x, la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it} \geq c$) por encima del eje de la x, y la proporción que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it} < c$) por debajo del eje de la x. La estimación DLLM subestima la movilidad ascendente y descendente en las cotas de ingresos bajos y la sobreestima en las cotas de ingresos altos. Sin embargo, la estimación de las transiciones de entrada y salida de la pobreza (movilidad en la línea de pobreza) son muy precisas.
Fuente: Estimación del autor con base en los datos de la CPS ASEC (panel 2005-2006).

Cuadro 1
Resultados de la regresión por MCO para el logaritmo del ingreso, PSID-CNEF 2005 y 2007

	2005	2007
Años de escolaridad	0.155*** (0.005)	0.157*** (0.006)
Edad	0.0425*** (0.0136)	0.0590*** (0.0157)
Edad al cuadrado	-0.000313* (0.000171)	-0.000538*** (0.000189)
Hombres	0.698*** (0.029)	0.593*** (0.032)
Negros	-0.362*** (0.027)	-0.429*** (0.030)
Hispanos	-0.579*** (0.084)	-0.435*** (0.093)
Constante	6.66*** (0.27)	6.45*** (0.33)
R Cuadrada	0.33	0.27
N	4,704	4,704

Notas: * p< 0.1, ** p< 0.05, *** p< 0.01. Errores estándar robustos entre paréntesis.
Fuente: Estimación propia con base en los datos del PSID-CNEF 2005 y 2007.

Este patrón no resulta único a los datos del PSID-CNEF. La Gráfica 3 muestra la curva de movilidad verdadera con intervalos de confianza y la estimación de DLLM para el panel de 1 año de los datos del CPS ASEC de uso público, para 2005 y 2006.²⁶ Analizamos la movilidad del ingreso total del hogar ajustado usando una escala de equivalencia de la raíz cuadrada de todos los hogares con jefes de hogar que tenían entre 25 y 54 años de edad en 2005. Los resultados de la regresión por MCO para los datos del CPS ASEC se muestran en el Cuadro 2. Los resultados de la curva de movilidad son muy similares a los que se obtienen con los datos del PSID, con el análisis DLLM subestimando la movilidad ascendente y descendente para cotas de ingresos muy bajos y sobreestimando ambos indicadores de movilidad en muchas cotas de ingresos medios y altos. En los datos del CPS ASEC, el análisis DLLM y las curvas de

26 Los intervalos de confianza estimados usando CPS ASEC replican las ponderaciones. Para una discusión de errores estándar y réplicas de ponderaciones en la CPS ASEC, ver el documento de us Census Bureau, «Estimating ASEC Variances with Replicate Weights», 2013. Las encuestas por hogar CPS ASEC se basan en las direcciones y no siguen a los individuos cuando se mudan, sino que en lugar de eso levantan la información de los nuevos residentes en el levantamiento posterior. Así, un procesamiento por *matcheo* debe ser hecho para construir un panel de la encuesta. Para una discusión detallada de la construcción de paneles de las CPS ASEC, ver B. Madrian y L. Lefgren, «An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents», *Journal of Economic and Social Measurement*, vol. 26, 2000, pp. 31–62. Como nuestra meta es únicamente validar técnicas de panel sintético, usamos un método simple para crear el panel. Incluimos hogares cuyo jefe de hogar es de la misma etnia y género en ambos años y cuya edad está entre un año menos y dos años más.

Cuadro 2

Resultados de la regresión por MCO para el logaritmo del ingreso, CPS ASEC (panel 2005-2006)

Escolaridad (excluyendo preparatoria)	2005	2006
Menos que preparatoria	-0.455*** (0.020)	-0.419*** (0.018)
Universidad incompleta	0.189*** (0.012)	0.186*** (0.012)
Universidad	0.534*** (0.013)	0.542*** (0.013)
Maestría	0.679*** (0.019)	0.692*** (0.019)
Doctorado o equivalente	0.941*** (0.030)	0.964*** (0.029)
Edad	0.005 (0.006)	0.003 (0.006)
Edad al cuadrado	-0.000097 (0.000073)	-0.000125* (0.000072)
Negros	-0.436*** (0.048)	-0.442*** (0.048)
Hispanos	-0.235*** (0.014)	-0.224*** (0.014)
Constante	10.071*** (0.115)	10.095*** (0.118)
R Cuadrada	0.21	0.21
N	22,156	22,156

Notas: * p < 0.1, ** p < 0.05, *** p < 0.01. Errores estándar robustos entre paréntesis.
Fuente: Estimación del autor con base en los datos de la CPS ASEC (panel 2005-2006).

movilidad verdadera convergen en las cotas de ingresos superiores, de modo que ninguno de los dos análisis resulta estadísticamente distinto, tanto para movilidad ascendente como descendente a partir del corte superior de 150 mil dólares o más.

Al utilizar datos de las encuestas NLSY para 1979 y 1997, también podemos probar la técnica de panel sintético con datos intergeneracionales. De nuevo usamos la escala de equivalencia de la raíz cuadrada para ajustar los ingresos del hogar. Medimos la movilidad de ingreso intergeneracional para la progeñe en sus años tempranos en la fuerza laboral (edades 26-30) en comparación con el ingreso de sus padres cuando los hijos eran adolescentes (14-18 en la muestra NLSY-1979 y 12-16 en la muestra NLSY-1997). En cada caso, promediamos el ingreso del hogar sobre dos años para reducir el sesgo de

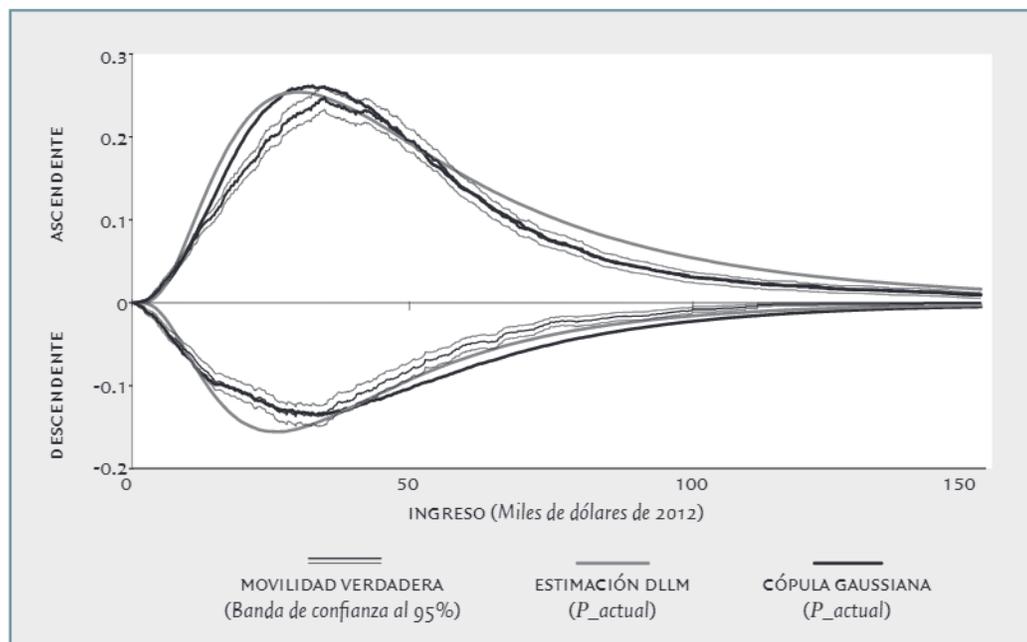
Cuadro 3

Resultados de la regresión por MCO para el logaritmo del ingreso, NLSY-1979

	Padres	Hijos
Años de escolaridad	0.063*** (0.002)	0.009*** (0.002)
Urbano	0.180*** (0.023)	0.137*** (0.028)
Regiones (excluyendo noreste)		
Norte	0.059** (0.028)	-0.186*** (0.035)
Sur	-0.047* (0.027)	-0.110*** (0.032)
Oeste	0.078*** (0.031)	-0.049 (0.037)
Raza/Etnicidad (excluyendo blancos)		
Negros	-0.639*** (0.026)	-0.548*** (0.030)
Hispanos	-0.403*** (0.029)	-0.372*** (0.033)
Otros no blancos	-0.448*** (0.025)	-0.158*** (0.037)
Edad	-0.006*** (0.002)	0.423* (0.217)
Edad al cuadrado	0.00011*** (0.00004)	-0.01260* (0.00675)
Industria del jefe del hogar (excluyendo ventas al por menor)		
Agricultura	-0.148** (0.067)	-0.141 (0.079)
Minería	-0.004 (0.370)	-0.090 (0.150)
Manufactura	-0.223* (0.100)	0.113** (0.049)
Transporte	-0.018 (0.053)	0.233*** (0.037)
Finanzas	-0.014 (0.130)	0.337*** (0.053)
Servicios empresariales	0.029 (0.099)	0.444*** (0.053)
Servicios personales	-0.099 (0.077)	0.064 (0.049)
Recreación	-0.201*** (0.046)	-0.223*** (0.061)
Servicios profesionales	-0.186* (0.099)	0.217 (0.117)
Otros	-0.246*** (0.051)	0.261*** (0.037)
Constante	9.54*** (0.05)	6.69*** (1.73)
R Cuadrada	0.33	0.21
N	5,005	5,005

Notas: * p < 0.1, ** p < 0.05, *** p < 0.01. Errores estándar robustos entre paréntesis.
Fuente: Estimación propia con base en los datos de la NLSY-1979.

Gráfica 4
 Movilidad intergeneracional de ingresos. Comparación de la movilidad verdadera con la estimación DLLM y la cópula Gaussiana con $\rho_{Actual} = 0.24$, NLSY-1979



Notas: La curva de movilidad verdadera, ascendente y descendente, con intervalos de confianza al 95%, y las estimaciones DLLM y de panel sintético usando cópula Gaussiana, utilizan la correlación conocida de los errores MCO: ρ_{Actual} . Para cada cota (c) sobre el eje de la x , la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it} \geq c$) por encima del eje de la x , y la proporción que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it} < c$) por debajo del eje de la x . La estimación DLLM subestima la movilidad ascendente y descendente en las cotas de ingresos bajos y la sobrestima en las cotas de ingresos altos. Sin embargo, la estimación de las transiciones de entrada y salida de la pobreza (movilidad en la línea de pobreza) son muy precisas.

Fuente: Estimación propia con base en los datos de la NLSY-1979.

atenuación del error de medición tanto como sea posible.²⁷ La regresión por MCO de la muestra NLSY-1979 para padres e hijos incluye edad, edad al cuadrado, años de escolaridad, *dummies* para el empleo primario del jefe del hogar cuando se encuentra en la categoría industrial, una *dummy* para estatus urbano, *dummies* de región y *dummies* para población de color no hispana, hispanos y otros no blancos. Los resultados de la regresión por MCO se muestran en el Cuadro 3.

La Gráfica 4 muestra la curva de movilidad verdadera y la estimación de panel sintético para varios valores de correlación. La correlación entre los errores del periodo 1 y 2 es aproximadamente de 0.24, valor que se usa para las estimaciones de panel sintético que se muestran en la gráfica.²⁸ Al comparar la estimación DLLM con la curva de movilidad verdadera, algunas cosas llaman la atención. Tanto para la movilidad ascendente, descendente, como para la movilidad intrageneracional —con base en los datos del PSID-CNEF y el CPS ASEC— la estimación DLLM excede la banda de

confianza del 95% de la curva de movilidad verdadera en muchas cotas de ingresos medios y altos. Aunque no se muestra en la figura, al asumir la normalidad bivariada, la estimación DLLM en realidad excede la estimación no paramétrica de su límite superior, mismo que indica que en algunas cotas, no hay una correlación entre los ingresos iniciales y finales ni en la movilidad ascendente ni en la descendente.

La Gráfica 5 muestra la movilidad intergeneracional y la estimación DLLM para la muestra NLSY-1997. Los resultados de la regresión por MCO para este grupo se encuentran en el Cuadro 4. En este caso, se usaron menos variables para mostrar cómo las estimaciones del panel sintético se comportan incluso con una R^2 relativamente baja en la regresión de mínimos cuadrados ordinarios.²⁹ La comparación entre la NLSY-1997 y las estimaciones DLLM y la curva de movilidad verdadera resulta muy similar a la comparación con la NLSY-1979. De nuevo, la estimación de movilidad DLLM excede la banda de confianza del 95% para la movilidad real en muchas cotas de ingresos medios y altos. En ambas muestras de la NLSY, las estimaciones de curvas de movilidad DLLM

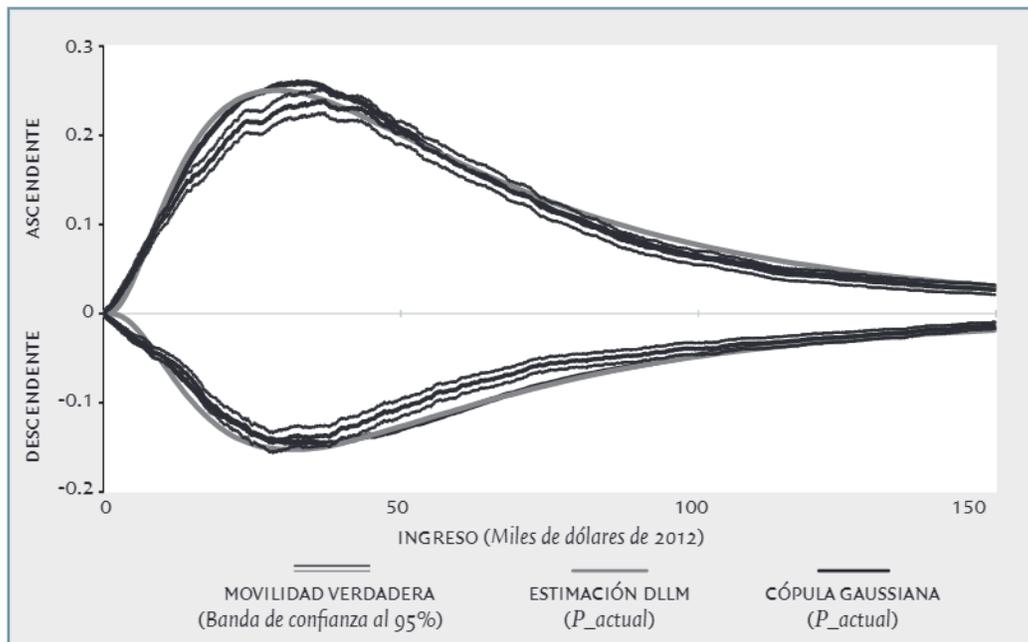
27 B. Mazumder, «Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data», *Review of Economics and Statistics*, vol. 87, núm. 2, 2005, pp. 235–255.

28 Las correlaciones verdaderas de error son de 0.241 para la muestra NLSY-1979 y 0.246 para la muestra NLSY-1997.

29 Cruces, et al., *op. cit.* muestran cómo incrementar el número de variables explicativas mejora la predicción y centra el rango entre los límites superior e inferior como ϵ_{it} da cuenta de una proporción más pequeña del ingreso del hogar.

Gráfica 5

Movilidad intergeneracional de ingresos. Comparación de la movilidad verdadera con la estimación DLLM y la cópula Gaussiana con $\rho_{Actual} = 0.24$, NLSY-1979



Notas: La curva de movilidad verdadera, ascendente y descendente, con intervalos de confianza al 95%, y las estimaciones DLLM y de panel sintético usando cópula Gaussiana, utilizan la correlación conocida de los errores MCO: ρ_{Actual} . Para cada cota (c) sobre el eje de la x , la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it} \geq c$) por encima del eje de la x , y la proporción que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it} < c$) por debajo del eje de la x . La estimación DLLM subestima la movilidad ascendente y descendente en las cotas de ingresos bajos y la sobrestima en las cotas de ingresos altos. Sin embargo, la estimación de las transiciones de entrada y salida de la pobreza (movilidad en la línea de pobreza) son muy precisas.

Fuente: Estimación propia con base en los datos de la NLSY-1997.

Cuadro 4
Resultados de la regresión por MCO para el logaritmo del ingreso, NLSY-1997

	Padres	Hijos
Años de escolaridad	0.066*** (0.003)	0.059*** (0.003)
Urbano	0.105*** (0.030)	-0.010 (0.032)
Regiones (excluyendo noreste)		
Norte	0.094** (0.042)	-0.114*** (0.043)
Sur	0.037 (0.040)	-0.039 (0.040)
Occidente	0.065 (0.044)	0.005 (0.044)
Raza/Etnicidad (excluyendo blancos)		
Negros	-0.797*** (0.034)	-0.625*** (0.033)
Hispanos	-0.599*** (0.038)	-0.221*** (0.036)
Otros no blancos	-0.123* (0.069)	-0.109 (0.069)
Constante	9.37*** (0.05)	9.75*** (0.06)
R Cuadrada	0.28	0.16
N	4,685	4,685

Notas: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Errores estándar robustos entre paréntesis.
Fuente: Estimación propia con base en los datos de la NLSY-1997.

para la movilidad ascendente y descendente exceden el límite superior no paramétrico sobre grandes rangos de cotas, por ejemplo, incluyendo casi todos los cortes de movilidad descendente por encima de 40 mil dólares en ambos casos.

4. Cópulas y dependencia en las distribuciones de errores no normales

En la sección anterior, mostramos que aun con la correlación conocida entre los ingresos del periodo 1 y el periodo 2, asumir que los errores tienen una distribución normal bivariada en las regresiones del logaritmo del ingreso puede arrojar estimaciones sesgadas de movilidad. Esas estimaciones pueden incluso exceder el límite superior no paramétrico, el cual no establece supuestos acerca de la distribución marginal de los términos de error. Por lo tanto, una fuente importante de este sesgo podría ser el supuesto de normalidad. DLLM prueban y rechazan el supuesto de normalidad en las distribuciones del error univariadas y bivariadas para ambos países analizados. Dang y Lanjouw, para una variedad de países, contraponen el logaritmo del ingreso y del consumo con distribuciones normales, y en todos los casos el error

también se desvía de la normalidad.³⁰ DLLM reconocen que «pese a este rechazo mantendremos el supuesto [...] y por lo tanto ilustramos el desempeño de nuestros métodos de límites paramétricos en una situación práctica típica donde el supuesto subyacente de las distribuciones no precisamente se sostiene».³¹

En este documento proponemos una técnica alternativa para utilizar la distribución observada de los errores con cópulas para estimar la distribución conjunta del ingreso en lugar de asumir normalidad bivariada. De este modo, podemos probar cómo el no suponer una distribución normal bivariada de los errores puede mejorar la estimación de las medidas de movilidad al usar paneles sintéticos. Las cópulas son funciones que relacionan distribuciones multivariadas con sus distribuciones marginales. Son especialmente útiles cuando las variables no son normales.³² Las cópulas se han usado en una amplia variedad de aplicaciones en economía, especialmente en finanzas, pero también para modelar ingreso y riqueza, incidencia de impuestos e inequidad, y la distribución del ingreso, desigualdad, y movilidad.³³

Para cada distribución conjunta de F existe una cópula C que relaciona la distribución conjunta con las distribuciones marginales F_1 y F_2 .³⁴

$$5.1 \quad F(x_1, x_2) = C(F_1^{-1}(x_1), F_2^{-1}(x_2)).$$

La elección de la cópula determina la estructura de dependencia entre los rangos en el periodo 1 y las distribuciones marginales del periodo 2.

30 Hai-Anh Dang y P. Lanjouw, *op. cit.*

31 En Hai-Anh Dang y P. Lanjouw, *op. cit.*, los autores intentan reducir la desviación de la normalidad al implementar una transformación Box-Cox de los ingresos de los periodos 1 y 2 para minimizar el sesgo de las distribuciones de ingreso transformadas. Esto reduce pero no elimina los sesgos reportados en la sección previa. Aunque con mucho menos frecuencia, el panel sintético generado con la transformación Box-Cox al p_{Actual} también excede la estimación no paramétrica DLLM del límite superior en algunas cotas. En ninguno de los conjuntos de datos analizados en este capítulo ocurre esto al usar la cópula Gaussiana.

32 P. Trivedi y D. Zimmer, «Copula Modeling: An Introduction for Practitioners», *Foundations and Trends in Econometrics*, vol. 1, núm. 1, 2005, pp. 1-111.

33 U. Cherubini, *et al.*, *Copula Methods in Finance*, West Sussex, John Wiley & Sons, 2004; M. Jäntti, *et al.*, «Modelling the Joint Distribution of Income and Wealth», *Working Paper*, International Association for Research in Income and Wealth, 2012; E. Bø, *et al.*, «Horizontal Inequity under a Dual Income Tax System: Principles and Measurement», *International Tax and Public Finance*, vol. 19, núm. 5, 2011, pp. 625-640; D. Zimmer y H. Youn Kim, «The Dependence Structure of Income Distribution», *Applied Economics*, vol. 44, núm. 27, 2012, pp. 3573-3583; A. Vinh, *et al.*, «Bivariate Income Distributions for Assessing Inequality and Poverty under Dependent Samples», *Economic Modelling*, vol. 27, núm. 6, 2010, pp. 1473-1483; S. Bonhomme y J. Robin, «Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000», *The Review of Economic Studies*, vol. 76, núm. 1, 2009, pp. 63-92.

34 P. Jaworski, *et al.* (eds.), *Copula Theory and Its Applications*, Nueva York, Springer, 2010.

Muchas clases de cópulas se han utilizado en la literatura. En este capítulo, usamos la cópula Gaussiana (Normal), la cual simula la estructura de dependencia de la distribución normal multivariada. La dependencia entre los rangos en la distribución del error será la misma que en el caso de la normalidad bivariada, pero la distribución marginal de F_1 y F_2 puede estimarse a partir de las distribuciones empíricas de los datos sin imponer normalidad como en DLLM. Hemos escogido usar la cópula Gaussiana en este capítulo por diversas razones. La primera es simplicidad. La cópula Gaussiana se determina por un parámetro simple, la correlación ρ entre los ingresos del periodo 1 y 2 al igual que los resultados en DLLM están bajo el supuesto de normalidad bivariada. De este modo, podemos comparar nuestros resultados con los suyos bajo el mismo parámetro de correlación para evaluar las diferentes técnicas. Otra ventaja de la cópula Gaussiana es que si el supuesto de normalidad bivariada DLLM se sostiene, la cópula y la estimación DLLM darán el mismo resultado.

Sin embargo, la cópula Gaussiana también tiene desventajas en relación con cópulas alternativas. Una deficiencia importante es que la cópula Gaussiana asume que no hay dependencia entre los valores extremos en las colas de la distribución.³⁵ Otras cópulas más flexibles, tales como la cópula t , aportan parámetros adicionales con los cuales determinar la dependencia de las colas. Sin embargo, el objetivo en este capítulo es mostrar cómo relajar el supuesto de normalidad usando cópulas puede ayudar en la estimación de la movilidad con paneles sintéticos, sin que esto signifique avalar el uso de un tipo de cópula en particular.³⁶

Generamos el panel sintético usando la cópula Gaussiana para estimar medidas de movilidad y entonces repetimos el proceso de manera aleatoria (*bootstrap*) para encontrar la curva de movilidad promedio. Para ello se echa mano del enfoque de cópula como sigue:

1. Correr la regresión por MCO de $\ln(y_{it})$ sobre las variables independientes x_{it} para conseguir $\hat{\epsilon}_{it}$ y $\hat{\epsilon}_{it}$ para cada individuo u hogar de las dos muestras.
2. Generar una matriz de dependencia del panel sintético r con n observaciones donde $r_t = (r_{1t}, \dots, r_{nt})$, $r = (r_1, r_2) \dots r_{jt} \in [0, 1]$ de la cópula Gaussiana con el parámetro de correlación ρ . Cada r_{jt} es el cuantil en

35 S. Demarta y A. McNeil, *op. cit.*

36 Dejamos para potencial trabajo futuro la tarea de seleccionar la cópula apropiada y estimar los parámetros adicionales de la cópula más allá de ρ .

la distribución del error para el individuo sintético j en el periodo t .

3. Para cada hogar, tomar una selección aleatoria de la distribución de la cópula del vector de dependencia r_j , donde $r_j = (r_{j1}, r_{j2})$.
4. Usar la densidad kernel para estimar la función de distribución acumulada empírica de la distribución del error del periodo 1 y el periodo 2, \hat{F}_1, \hat{F}_2 . Entonces para cada hogar i con un error j del panel sintético, los errores del panel sintético en el periodo 1 y el periodo 2 son $\tilde{\epsilon}_{i1} = \hat{F}_1^{-1}(r_{j1})$ y $\tilde{\epsilon}_{i2} = \hat{F}_2^{-1}(r_{j2})$.³⁷
5. Repetir para todos los individuos i para conseguir los valores completos del ingreso del panel sintético con $\tilde{y}_{i1} = x_{i1}'\beta_1 + \tilde{\epsilon}_{i1}$ y $\tilde{y}_{i2} = x_{i2}'\beta_2 + \tilde{\epsilon}_{i2}$, donde x_{i1}^2 son las covariables del periodo 1 para cada individuo i observado en el periodo 2, y construir la curva de movilidad de los ingresos sintéticos.
6. Repetir pasos 1-5 con B réplicas y promediar la curva de movilidad sobre todas las réplicas para estimar el panel sintético promedio de la curva de movilidad (como en DLLM para la estimación del límite superior de movilidad).

El próximo paso es validar la técnica de cópula con los datos existentes. La Gráfica 2 y Gráfica 3 también muestran la estimación de cópula de las curvas de movilidad con el valor verdadero de ρ , de la muestra PSID-CNEF y el panel de 1 año para la muestra CPS ASEC, respectivamente. Ambas técnicas aportan una estimación razonablemente precisa de la movilidad experimentada, especialmente cerca de la línea de pobreza (aproximadamente \$11,000 en cada una).³⁸

Como estamos estudiando movilidad intergeneracional, analizamos nuevamente las muestras de 1979 y 1997 de la NLSY y comparamos los resultados DLLM con la cópula Gaussiana. En ambos casos, calculamos la estimación de la cópula de movilidad del panel sintético usan-

37 Si las observaciones son igualmente ponderadas, los errores del periodo 1 pueden ser extraídos directamente de los valores de los errores observados. Sin embargo, si las observaciones tienen ponderaciones, se debe estimar la función de distribución acumulada empírica y tomar los errores de la distribución estimada.

38 En el Apéndice 2, disponible en línea en www.ceey.org.mx, comparamos las desviaciones absolutas y cuadradas de las estimaciones del panel sintético de la movilidad verdadera observada en el panel, incluyendo la descomposición por cuartiles. Estas figuras (A2.1-4 para la muestra PSID-CNEF y A2.5-8 para la muestra CPS ASEC) grafican la desviación para cada cota y la desviación media arriba de cada cota. La desviación en cada cota se mide para $k=1,2$ (absoluta, al cuadrado) como $d_k(c) = |m_{D, Verdadera}(y, c) - m_{D, Sintetica}(y, c)|^k + |m_{D, Verdadera}(y, c) - m_{D, Sintetica}(y, c)|^k$. La desviación media arriba de cada cota (c) es el promedio de las desviaciones mayores o igual a c tal que $\mu_k^+(c) = \frac{1}{n} \sum_{i=c}^n |m_{D, Verdadera}(y, c) - m_{D, Sintetica}(y, c)|^k + |m_{D, Verdadera}(y, c) - m_{D, Sintetica}(y, c)|^k$. Mientras las desviaciones absolutas y cuadradas y sus medias para las estimaciones de cópula no son siempre menores que las estimaciones DLLM, la cópula está menos sesgada en la mayoría de las cotas en ambas muestras, PSID-CNEF y CPS ASEC. En las descomposiciones por cuartiles, la cópula también supera a DLLM en casi todas las cotas.

do la verdadera correlación del error en los periodos 1 y 2 ($\rho_{Actual} = 0.24$). Los resultados para la muestra NLSY-1979 se muestran en la Gráfica 4 y para la muestra NLSY-1997 en la Gráfica 5.³⁹

La Gráfica 6 muestra los resultados completos de la curva de movilidad incluyendo la descomposición por cuartiles para la muestra NLSY-1979. Tanto la cópula como DLLM sobrestiman la movilidad ascendente en el primer cuartil (DLLM a niveles más bajos de ingresos). En el segundo cuartil, ambas estimaciones de panel sintético tienen las curvas de movilidad ascendente y descendente desplazadas a la izquierda, pero en ambos casos, la cópula se encuentra más cerca al valor verdadero que el valor DLLM. Los resultados de cópula coinciden muy bien con los valores verdaderos en el tercer cuartil, y tanto las cópulas como DLLM sobrestiman la movilidad descendente en el cuarto cuartil. Para el caso de la NLSY-1997, los resultados son casi idénticos (y aquí no se reportan).

Una razón central por la que los resultados DLLM puedan estar sesgados es que las distribuciones predichas del panel sintético, inicial y final, no coinciden con los datos, debido al supuesto de normalidad. Como resultado de lo anterior, las curvas de movilidad DLLM en los cuartiles 2 y 3 están desplazadas a la izquierda. La cópula, al utilizar las distribuciones empíricas del error, no sufre tanto de esta limitación. En estas figuras queda claro que ni la cópula ni las estimaciones DLLM para la movilidad son perfectas. Sin embargo, las comparaciones de la desviación muestran que la cópula generalmente ofrece menos sesgo en la estimación de la movilidad verdadera que DLLM al usar estos conjuntos de datos.⁴⁰

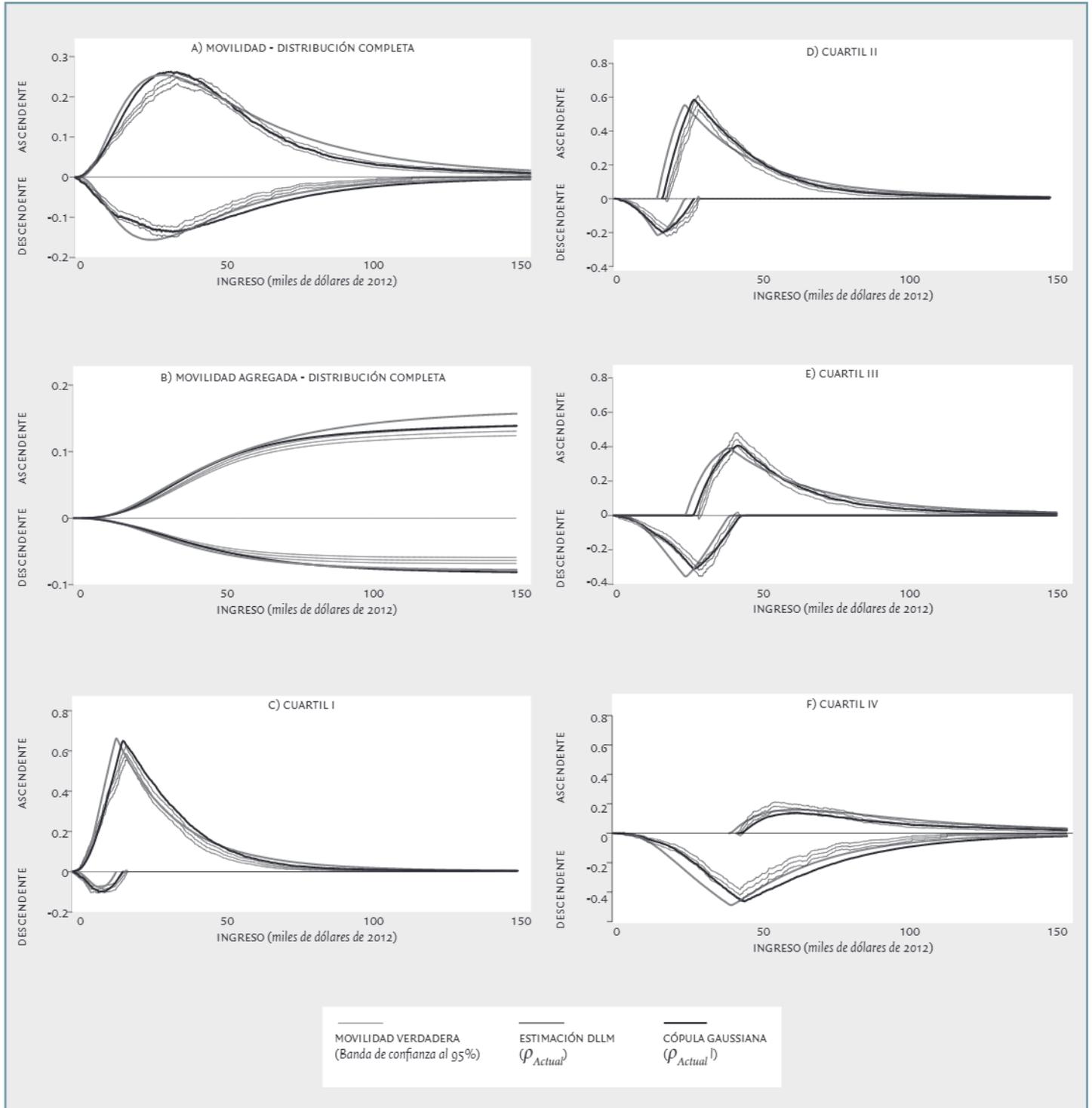
También calculamos las estimaciones del panel verdadero y sintético para distintas medidas de movilidad, incluyendo la elasticidad intergeneracional del ingreso (IGE), la correlación entre logaritmos de ingreso, y matrices de transición por quintiles. Estas estimaciones se muestran en el Apéndice 1. Para cada medida de movilidad, los resultados se reportan tanto para ρ_{Actual} como para un rango de correlaciones posibles $\rho_L < \rho_{Actual} < \rho_H$. En cada uno de los cuatro conjuntos de datos que hemos discutido, la estimación de cópula de la

39 En las figuras de las desviaciones (A2.9-12 para los datos NLSY-1979, y A2.13-16 para los datos NLSY-1997), para casi cada cota, la cópula supera a DLLM tanto al nivel agregado (la curva de movilidad m y el agregado de la curva de movilidad M) como para cada cuartil en la descomposición. Ver Apéndice 2, disponible en www.ceey.org.mx.

40 Para ver las comparaciones ir a Apéndice 2 disponible en línea en www.ceey.org.mx

Gráfica 6

Resultados de la curva de movilidad completa y descomposición por cuartiles con datos de la NLSY-1979. Comparación de la movilidad verdadera con la estimación DLLM y la cópula Gaussiana con $\rho_{Actual} = 0.24$.



Notas: La curva de movilidad verdadera, ascendente y descendente, y la descomposición por cuantiles de la NLSY-1979 con intervalos de confianza al 95%, y las estimaciones de panel sintético DLLM y cópula Gaussiana utilizan la correlación conocida de los errores MCO: ρ_{Actual} . Para cada cuota (c) sobre el eje de la x, la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{1t} < c$ y $y_{2t} \geq c$) por encima del eje de la x y la proporción que experimenta movilidad descendente ($y_{1t} \geq c$ y $y_{2t} < c$) por debajo del eje x.

Fuente: Estimación propia con base en los datos de la NLSY-1979.

elasticidad intergeneracional del ingreso (IGE) (Cuadro A1.1) y la correlación del logaritmo del ingreso (Cuadro A1.2) con ρ_{Actual} está más cerca al valor verdadero que DLLM. En las dos muestras NLSY, la cópula IGE y las estimaciones de correlación son casi las mismas que los valores verdaderos.

Las dos técnicas de panel sintético tienen casi idénticos resultados al estimar las matrices de transición por quintiles (Cuadro A1.3). Tanto las técnicas DLLM como la cópula fueron más precisas en la estimación de las transiciones intergeneracionales en las muestras NLSY que en las intrageneracionales en las muestras del PSID-CNEF y el CPS ASEC. Para la mayoría de las celdas en las matrices de transición por quintiles de las muestras NLSY, la probabilidad de transición verdadera se encuentra dentro del intervalo de confianza al 95% de las estimaciones DLLM y cópula. El hecho de que la cópula ofrezca poca o ninguna ventaja al estimar transiciones, apoya la noción de que los sesgos en los paneles sintéticos DLLM se deben a los supuestos de la distribución, tal que estos supuestos son probablemente menos relevantes al analizar cambios en rangos en contraposición con cambios en ingresos.

Los resultados de la validación de la cópula Gaussiana nos dan confianza en la aplicabilidad de esta técnica para estimar movilidad intergeneracional con paneles sintéticos. También sugiere que una avenida para investigación futura es probar el uso de cópulas con mayor dependencia en las colas de la distribución del error. Esto puede reducir los sesgos causados por la sobrestimación de la movilidad descendente en la cola derecha de la distribución y de la movilidad ascendente en la cola izquierda de la distribución.

5. La EMOVI y la movilidad intergeneracional en México

Ahora vamos a aplicar esta técnica para estimar movilidad intergeneracional en México, donde no existen datos de panel para hacerlo. Para construir esta estimación usamos dos fuentes de datos distintos. Una debe incluir información acerca del ingreso (y_{i2}) y características de los hogares de los hijos (x_{i2}^2). Esta fuente de datos también incluye información retrospectiva acerca de los padres de la cohorte de los hijos (x^2). La segunda fuente de datos debe incluir ingreso (y_{i1}) y características (x_{i1}) para un corte transversal de individuos de la cohorte de los padres. Las variables x_{it} en el conjunto de los datos

de los padres deben corresponder a la información retrospectiva x_{it}^2 de la cohorte de los hijos.

Para los datos de la cohorte de los hijos y los datos retrospectivos, usamos la Encuesta ESRU de movilidad social en México (EMOVI). Esta encuesta se realizó sobre cortes transversales de los hogares mexicanos en 2006 y 2011 con jefes de hogar entre las edades de 25 y 64 años. La EMOVI-2006 incluye 4,743 hogares que reportan ingreso positivo e información de los bienes de los padres. La EMOVI-2011 incluye 3,818 de individuos con estas mismas características. La encuesta contiene diversas preguntas acerca de las características del hogar y de los individuos en la generación actual, lo que incluye bienes del hogar, ingreso mensual, ocupación y educación. Crucial para este estudio, las encuestas EMOVI también incluyen una serie de preguntas acerca de las características de sus padres y los hogares de la infancia de los individuos entrevistados. Se pide a los informantes recordar qué ocurría en sus hogares a la edad de 14 años, e incluyen información acerca de las ocupaciones de los padres, niveles de educación, localidad y características del hogar. Las preguntas retrospectivas se centran en las características que un individuo probablemente recuerda con razonable precisión. Por ejemplo, no se pregunta a los hijos acerca del ingreso laboral mensual, sino que se les consulta la ocupación de sus padres. A los hijos también se les hizo una serie de preguntas retrospectivas acerca de las características del hogar que son buenos predictores del ingreso, tales como propiedad de automóviles, teléfono, televisión, acceso a electricidad y agua entubada dentro de la vivienda. La técnica de panel sintético requiere variables x_{it} que sean buenos predictores del ingreso. Se requiere que la relación sea causal, de modo que estas preguntas acerca de bienes del hogar y acceso a servicios son especialmente valiosas al generar estimaciones precisas de movilidad.

Usamos la Encuesta Nacional de Ingresos y Gastos del Hogar (ENIGH) como la fuente de datos sobre las características del ingreso y del hogar de la cohorte de los padres. La ENIGH es una encuesta representativa para los hogares en México. Incluye información sobre ingreso, consumo y varias características del hogar. La ENIGH se levantó en 1984, 1989 y cada dos años a partir de 1992. Las encuestas EMOVI y ENIGH son compatibles para este análisis debido a que hay una cantidad significativa de tipo de información que se empalma entre las preguntas retrospectivas de las encuestas EMOVI y las preguntas contemporáneas en la ENIGH.

Emparejamos un subconjunto de los individuos de las encuestas EMOVI con individuos de las cohortes de sus padres en la ENIGH. Debido a que las preguntas retrospectivas en la encuesta EMOVI preguntan por las características de los padres y los hogares cuando los informantes tenían 14 años de edad, construimos los paneles sintéticos para un subconjunto de los individuos en las muestras EMOVI. Individuos con 36 años en 2006 contaban con 14 años en 1984, el primer año en que se levantó la encuesta ENIGH. Así restringimos nuestro foco a aquellos individuos de la EMOVI-2006 que tenían entre 30 y 39 (nacidos entre 1966 y 1976) y los emparejamos con individuos en la cohorte de sus padres en la ENIGH-1984. Los individuos que tenían 36 años de edad en la EMOVI-2011, tenían 14 años en 1989. Como resultado, usamos la ENIGH-1989 como la cohorte emparejada a los padres de los individuos de la EMOVI-2011, para aquéllos con edades entre 30 y 39 (nacidos entre 1971 y 1981). El ingreso del hogar que se usa es el ingreso del hogar equivalente al individuo. Para ello se usa la raíz cuadrada de la escala de equivalencia. En ambas encuestas ENIGH, incluimos sólo a los hogares con hijos en la muestra para emparejarlos con las cohortes de los padres de las encuestas EMOVI.

Para que las variables se incluyan en x_{11} , x_{12} y x_{11}^2 , restringimos nuestra atención a las variables en tres secciones de la encuesta, la encuesta de la cohorte de los padres (ENIGH), la encuesta de los hogares de los hijos adultos (EMOVI) y la encuesta de los hogares de los hijos acerca de las características de los hogares de sus padres (preguntas retrospectivas EMOVI). Esto incluye nivel educativo, ocupación, región, propiedad de automóvil, teléfono, televisión, acceso a agua entubada dentro del hogar y electricidad. Las regresiones también incluyen *dummies* para varios tamaños de ciudad. Para ello se echa mano de datos del censo de 1980 para la ENIGH-1984 y la cohorte de los padres de la EMOVI-2006, del censo de 1990 para la ENIGH-1989 y la cohorte de los padres de la EMOVI-2011, y del conteo de población y vivienda de 2005 para las cohortes de los hijos de las EMOVI 2006 y 2011. Nuestra muestra de los jefes de hogar de 30-39 años es de 1,388 casos para la EMOVI-2006 y 1,042 para la EMOVI-2011.

El Cuadro 5 muestra los estadísticos de resumen para las variables de regresión en las cohortes de padres de la ENIGH y la EMOVI. Si la memoria no falló y las encuestas ENIGH fueron una muestra representativa de los padres de las cohortes EMOVI, las características de los

Cuadro 5
Estadísticos de resumen para la ENIGH y la cohorte de los padres en la EMOVI

	ENIGH 1984	EMOVI-2006 padres	ENIGH 1989	EMOVI-2011 padres
Edad (en el año de la encuesta del padre)	40.4 (12.7)	43.6 (9.8)	40.4 (12.7)	47.9 (9.4)
Escolaridad (excluyendo educación primaria)				
Ninguna o primaria incompleta	0.551	0.560	0.655	0.443
Secundaria	0.131	0.079	0.131	0.211
Universidad	0.055	0.027	0.004	0.058
Ocupación (excluyendo ventas al por menor)				
Profesionales	0.016	0.013	0.028	0.027
Técnicos	0.023	0.005	0.031	0.027
Educación	0.027	0.011	0.029	0.008
Entretenimiento	0.010	0.008	0.008	0.002
Directores	0.018	0.001	0.022	0.001
Agricultura	0.319	0.325	0.263	0.209
Directores industriales	0.022	0.007	0.025	0.009
Trabajadores industriales	0.226	0.250	0.224	0.261
Administrativo	0.050	0.014	0.050	0.010
Ventas al por menor (ambulatorio)	0.011	0.011	0.019	0.012
Servicios	0.044	0.027	0.046	0.034
Servicio doméstico	0.013	0.018	0.011	0.000
Transporte	0.051	0.130	0.058	0.092
Región (excluyendo el norte)				
Centro	0.412	0.332	0.356	0.274
Oeste	0.117	0.143	0.128	0.056
Este	0.141	0.160	0.140	0.103
Sur	0.102	0.138	0.168	0.118
Tamaño de la ciudad (Grande, > 100,000 excluidas)				
Muy pequeñas ($\leq 10,000$)	0.099	0.118	0.057	0.065
Pequeñas (10,001-35,000)	0.238	0.220	0.168	0.156
Medianas (35,001-100,000)	0.228	0.174	0.221	0.187
Cuenta con automóvil	0.146	0.204	0.170	0.234
Agua entubada dentro del hogar	0.616	0.491	0.704	0.580
Electricidad en el hogar	0.872	0.717	0.884	0.906
Cuenta con teléfono	0.143	0.183	0.152	0.162
Cuenta con televisión	0.674	0.532	0.783	0.807
N	3,076	1,388	7,307	1,042

Notas: Desviación estándar para la edad del jefe del hogar entre paréntesis. Las demás variables son binarias y la desviación estándar es igual a $(p(1-p))^{1/2}$.

Fuente: Estimación propia con base en los datos de la ENIGH-1984 y 1989 y la EMOVI-2006 y 2011.

padres deben ser las mismas entre cada par de la cohorte de los padres de la ENIGH y la EMOVI. Para ambos pares (ENIGH-84-EMOVI-06 y ENIGH-89-EMOVI-11) hay algunas variables con diferencias relativamente grandes en las medias, incluyendo, por ejemplo, las *dummies* para educación secundaria y agua entubada dentro del hogar. Sin embargo, en general, las medias de las

variables incluidas para cada par de la cohorte de los padres se parecen. Los resultados de la regresión para las cohortes padre-hijo, calculados para la ENIGH-1984 con la EMOVI-2006, y la ENIGH-1989 con la EMOVI-2011, se presentan en el Cuadro 6. Para ambas cohortes de padres, la R^2 es igual a 0.50; para las cohortes de los hijos, especialmente en la EMOVI-2011, la R^2 es más baja. Esto

Cuadro 6

Resultados de la regresión por MCO para el logaritmo del ingreso, ENIGH y EMOVI

	ENIGH 1984 / EMOVI 2006		ENIGH 1989 / EMOVI 2011	
	padres (1984)	hijos (2006)	padres (1989)	hijos (2011)
Escolaridad	0.0052 (0.0043)	0.0318 (0.1455)	0.0049 (0.0030)	-0.0239 (0.1838)
Escolaridad al cuadrado	0.000021 (0.000046)	0.000571 (0.002110)	0.000020 (0.000031)	0.000276 (0.002688)
Escolaridad (educación primaria excluida)				
Sin escolaridad / Primaria incompleta	-0.198*** (0.027)	-0.151*** (0.056)	-0.170*** (0.020)	-0.025 (0.079)
Secundaria	0.106*** (0.035)	0.083** (0.037)	0.177*** (0.027)	0.169*** (0.052)
Universidad	0.331*** (0.062)	0.647*** (0.077)	0.575*** (0.108)	0.544*** (0.098)
Ocupación (excluyendo ventas al por menor)				
Profesionales	0.116 (0.092)	0.230* (0.114)	0.252*** (0.051)	0.407*** (0.121)
Técnicos	0.030 (0.071)	0.184* (0.103)	0.045 (0.043)	0.331 (0.165)
Educación	0.008 (0.076)	0.287* (0.137)	0.113*** (0.045)	0.165 (0.168)
Entretenimiento	-0.108 (0.102)	0.173 (0.179)	0.185** (0.074)	-0.016 (0.317)
Directores	0.447*** (0.081)	0.229 (0.179)	0.460*** (0.053)	0.618*** (0.226)
Agricultura	-0.132*** (0.034)	-0.329*** (0.070)	-0.112*** (0.024)	-0.079 (0.093)
Directores industriales	0.185*** (0.071)	0.208* (0.132)	0.269*** (0.050)	0.176 (0.165)
Trabajadores industriales	-0.082*** (0.032)	0.069* (0.041)	-0.018 (0.022)	0.099* (0.058)
Administrativos	0.053 (0.051)	0.083 (0.077)	0.151*** (0.036)	0.192*** (0.086)
Ventas al por menor (ambulatorios)	-0.095 (0.083)	-0.075 (0.135)	-0.079 (0.056)	-0.209* (0.135)
Servicios	-0.076* (0.049)	-0.067 (0.072)	0.015 (0.037)	0.157** (0.087)
Servicio doméstico	-0.414*** (0.095)	-0.317*** (0.110)	-0.002 (0.062)	-0.157 (0.119)
Transporte	0.010 (0.047)	-0.069 (0.067)	0.064** (0.033)	0.087 (0.213)

	ENIGH 1984 / EMOVI 2006		ENIGH 1989 / EMOVI 2011	
	padres (1984)	hijos (2006)	padres (1989)	hijos (2011)
Región (norte excluido)				
Centro	-0.145*** (0.026)	-0.159*** (0.042)	-0.166*** (0.019)	-0.094* (0.061)
Oeste	-0.138*** (0.033)	0.033 (0.044)	-0.133*** (0.022)	-0.233*** (0.074)
Este	-0.181*** (0.032)	-0.225*** (0.067)	-0.211*** (0.023)	-0.152** (0.068)
Sur	-0.045 (0.032)	-0.380*** (0.069)	-0.216*** (0.024)	-0.226*** (0.067)
Tamaño de ciudad (Grande, > 100 mil excluidas)				
Muy pequeñas (≤10 mil)	-0.127*** (0.041)	-0.114 (0.078)	-0.302*** (0.033)	-0.114 (0.102)
Pequeñas (10 mi l - 35 mil)	-0.192*** (0.028)	-0.210*** (0.049)	-0.217*** (0.022)	-0.124** (0.059)
Medianas (35 mil - 100 mil)	-0.035 (0.027)	-0.053 (0.053)	-0.146*** (0.021)	-0.195*** (0.051)
Cuenta con automóvil	0.387*** (0.032)	0.232*** (0.036)	0.365*** (0.021)	0.284*** (0.044)
Agua entubada dentro del hogar	0.200*** (0.024)	0.279*** (0.051)	0.180*** (0.018)	0.226*** (0.053)
Electricidad en el hogar	0.045 (0.035)	0.379** (0.132)	0.126*** (0.025)	-0.141 (0.141)
Cuenta con teléfono	0.256*** (0.033)	0.157*** (0.036)	0.362*** (0.021)	0.022 (0.047)
Cuenta con televisión	0.273*** (0.027)	0.239*** (0.071)	0.302*** (0.021)	0.140 (0.132)
Constante	7.40*** (0.10)	6.5*** (2.50)	7.2*** (0.07)	7.9** (3.12)
R Cuadrada	0.50	0.41	0.50	0.27
N	3,076	1,388	7,307	1,042

Notas: *p < 0.1, **p < 0.05, ***p < 0.01. Errores estándar robustos entre paréntesis.
Fuente: Estimación propia con base en los datos de la ENIGH 1984 y 1989 y la EMOVI 2006 y 2011.

puede deberse en parte al más bajo poder predictivo de algunas variables sobre el ingreso en la década de 2000, comparado con los ochenta, tales como contar con una televisión, tener electricidad, etc.⁴¹

A fin de construir los paneles sintéticos, debemos estimar primero la correlación del error MCO (ρ) para cada periodo. Para lo anterior seguimos a Dang y Lanjouw en la estimación de ρ_{y1y2} usando grupos de cohorte.⁴² En nuestro caso, construimos cohortes de las características del padre de las muestras ENIGH 1984 y 1989 y las características del padre de las muestras EMOVI 2006 y 2011. Usamos cohortes de edad-tamaño de ciudad para las generaciones de los padres. Creamos cohortes de edad en grupos de tres años para quienes tenían de 35 a 52 años, y también para padres que eran menores de 35 años y mayores de 52 años en 1984 y 1989, respectivamente, con cada cohorte de edad dividida por las 4 *dummies* de tamaño de ciudad tal como se representa en el Cuadro 5. Esto permite generar 32 grupos de cohortes. Al usar la ecuación (2.4) calculamos ρ_{y1y2} sobre las muestras conjuntas de las ENIGH-1984/1989 y las EMOVI-2006/2011. El propósito fue incrementar el tamaño de los grupos de cohortes. Esto proporcionó una correlación estimada del logaritmo del ingreso de 0.39, la cual, aunque poco más alta, fue similar a la correlación en las muestras NLSY ($\rho_{y1y2}^{NLSY79}=0.36$ y $\rho_{y1y2}^{NLSY97}=0.31$). Entonces estimamos la correlación residual de MCO al usar la ecuación (2.7). Utilizamos las x_i de las preguntas retrospectivas de la EMOVI, lo que resulta en $\rho_{Estimada}^{84-06}=0.12$ y $\rho_{Estimada}^{89-11}=0.18$. La ρ estimada más alta para la ENIGH-1989-EMOVI-2011 se debe a la R^2 más baja en la regresión del ingreso de los hijos de la EMOVI-2011.

La Gráfica 7 muestra la curva estimada de movilidad intergeneracional para los periodos 1984-2006 y 1989-2011 al utilizar las correlaciones estimadas así como las estimaciones del umbral superior e inferior con $\rho_U=0.05$ y $\rho_L=0.30$.

Los umbrales superior e inferior aportan un rango muy estrecho de estimaciones de movilidad, con una diferencia máxima entre los dos en ambos periodos para

cualquier cota de 0.024. Debido a los estrechos límites sobre las estimaciones de movilidad, procedemos con nuestro análisis usando $\rho_{Estimada}^{43}$. Con $\rho_{Estimada}^r$ encontramos que ambos tipos de movilidad, la ascendente y la descendente, fueron mayores en los niveles de ingresos más bajos para la cohorte 1984-2006 y mayores en los niveles de ingresos más altos para la cohorte 1989-2011. La Gráfica 7 también muestra la movilidad en descomposición por cuantiles para ambos periodos, 1984-2006 y 1989-2011. La brecha entre las curvas de movilidad de las dos cohortes se debe principalmente a las diferencias en la movilidad de los hijos en los cuantiles de mayores ingresos.

Sin embargo, estas diferencias absolutas en la movilidad se deben principalmente a las diferencias en las distribuciones del ingreso inicial y final, para ambas cohortes de padres e hijos en las ENIGH 1984 y 1989, así como en las EMOVI 2006 y 2011. La Gráfica 8 muestra la curva de movilidad y la descomposición para la movilidad de rango. Prácticamente no hay diferencias en la movilidad de rango entre las dos cohortes ENIGH-EMOVI en cualquiera de los cuantiles para cada cota. Esta persistencia de la movilidad de rango sobre lapsos cortos resulta consistente con los datos de la NLSY y con otro trabajo sobre movilidad en el tiempo en los Estados Unidos, de Chetty, *et al.*, que se elabora con base en datos de impuestos.⁴⁴

Debido a la estabilidad en las curvas de movilidad de rango y con el fin de comparar nuestros resultados con otras estimaciones de movilidad intergeneracional en México y en otros lugares, también observamos las matrices de transición por quintiles y por deciles, tal como se muestra en los cuadros 7 y 8. Al usar la EMOVI 2011 para personas ocupadas entre 30 y 50 años de edad (y no sólo para individuos entre 30-39 años como en este estudio), Vélez, Campos y Huerta encontraron que 48% de los individuos que crecieron en hogares en el quintil más bajo, permanecieron ahí en la edad adulta. Además, 52% de los hijos que crecieron en hogares en el quintil de estatus más alto, tampoco se movieron.⁴⁵

41 Debido a la más baja R^2 de la regresión del hijo, también probamos una variación del procedimiento de cópula. En esta variación, los errores de los hijos ϵ_{it} no fueron remplazados por una selección aleatoria de la distribución empírica basada en la matriz de dependencia r . En lugar de eso, cada posición del individuo en la distribución del error fue emparejada a la posición correspondiente en la matriz de dependencia donde $r_{ij}=r_{ji}$. El residual asignado a los padres de cada individuo fueron entonces $\epsilon_{it} = F^{-1}(r_{ij})$, del periodo correspondiente al rango 1 de la matriz de dependencia. Los resultados de movilidad fueron virtualmente idénticos para los dos métodos para todos los periodos, tanto para los datos mexicanos como para los de la muestra NLSY de Estados Unidos.

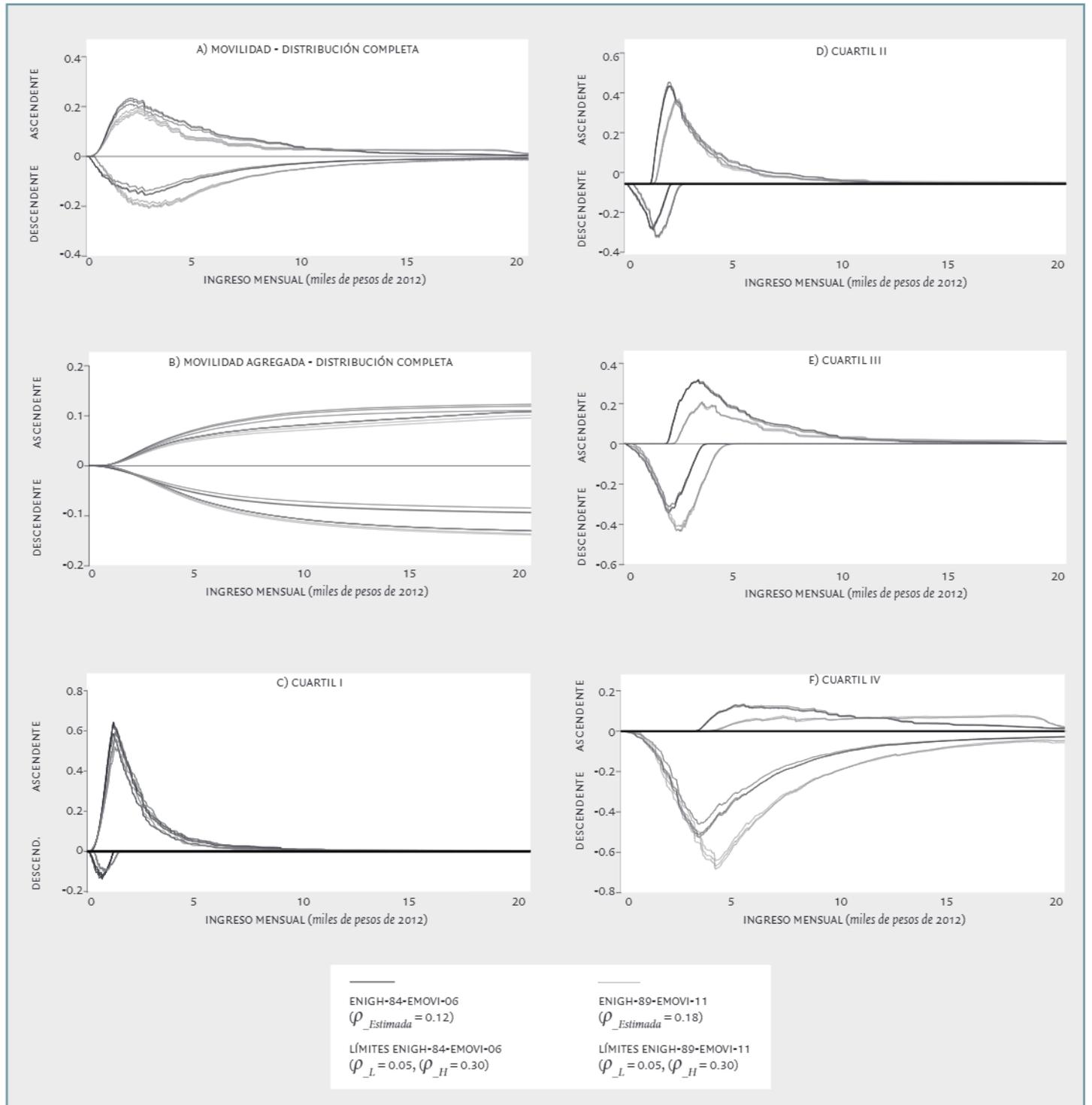
42 Hai-Anh Dang y P. Lanjouw, *op. cit.*

43 Las estimaciones de las medidas de movilidad usando ρ_L y ρ_M son reportadas en el Apéndice 1.

44 R. Chetty, *et al.*, «Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility», NBER Working Paper 19844, National Bureau of Economic Research, 2014.

45 R. Vélez-Grajales, *et al.*, *Informe Movilidad Social en México 2013. Imagina tu futuro*, México, Centro de Estudios Espinosa Yglesias, 2013.

Gráfica 7
 Estimación de la curva de movilidad y descomposición por cuartiles de la movilidad intergeneracional
 en México usando cópulas con límites

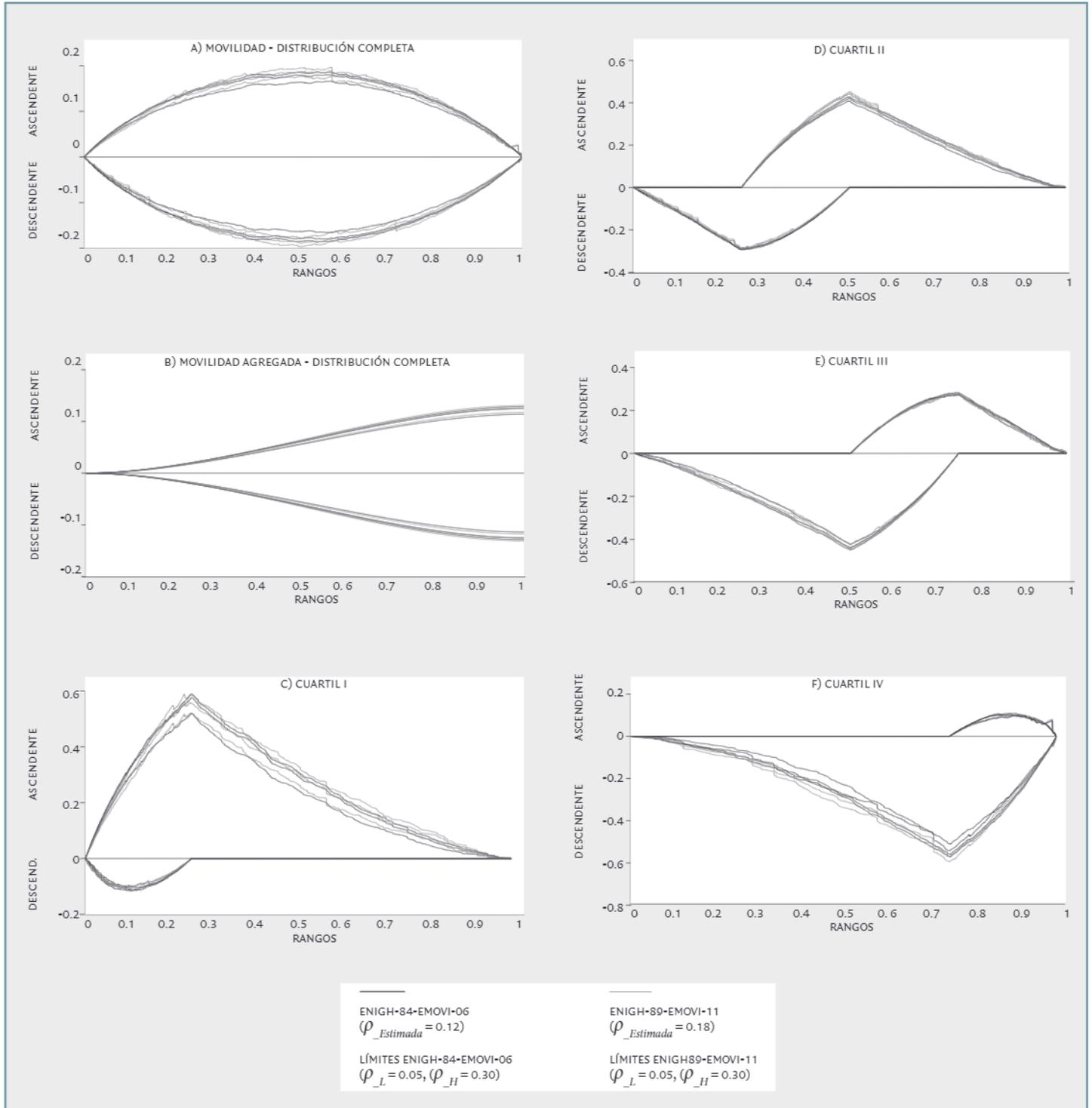


Notas: Las estimaciones de cópula de la curva de movilidad, ascendente y descendente, y la descomposición por cuartiles para la ENIGH-84-EMOVI-06 (nacidos en 1966-1976) y la ENIGH-89-EMOVI-11 (nacidos en 1971-1981) utilizan la correlación estimada de los errores MCO: $\rho_{Estimada}$ y los límites inferior y superior. Para cada cota (c) sobre el eje de la x, la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente ($y_{it} < c$ y $y_{it+1} \geq c$) por encima del eje de la x y la proporción que experimenta movilidad descendente ($y_{it} \geq c$ y $y_{it+1} < c$) por debajo del eje de la x.

Fuente: Estimación propia con base en los datos de la EMOVI-2006 y 2011 y de la ENIGH-1984 y 1989.

Gráfica 8

Estimación de la curva de movilidad por rangos y descomposición por cuartiles de la movilidad intergeneracional en México usando cópulas con límites



Notas: Las estimaciones de cópula de la curva de movilidad por rangos, ascendente y descendente, y la descomposición por cuartiles para la ENIGH84-EMOVI06 (nacidos en 1966-1976) y la ENIGH89-EMOVI11 (nacidos en 1971-1981) utilizan la correlación estimada de los errores MCO: $\rho_{Estimada}$ y los límites inferior y superior. Para cada cota (c) sobre el eje de la x, la curva de movilidad muestra la proporción de la población que experimenta movilidad ascendente $F_1^{-1}(y_H) \geq c$ y $F_2^{-1}(y_D) \geq c$ por encima del eje de la x y la proporción que experimenta movilidad descendente $F_1^{-1}(y_H) \geq c$ y $F_2^{-1}(y_D) < c$ por debajo del eje de la x.

Fuente: Estimación propia con base en los datos de la EMOVI 2006 y 2011 y de la ENIGH 1984 y 1989.

Al usar ingreso y sólo los hogares de la EMOVI cuyos jefes de hogar oscilan entre los 30 y 39 años, estimamos que en las dos cohortes 35-38% de los hijos que crecieron en los hogares del quintil más pobre se quedaron ahí, y 37-39% de los hijos que nacieron en hogares del quintil más alto también se mantuvieron en ese grupo. Además, los hijos que nacieron en hogares del quintil más pobre tienen una probabilidad cinco veces mayor de finalizar en el quintil más pobre que los hijos que nacieron en el quintil más rico.

Esto resulta similar, en lo general, a los datos de los Estados Unidos. 43% de los hijos que nacen en el quintil más pobre se quedaron ahí para la NLSY-1979, y 40% lo hicieron para la NLSY-1997. En ambos casos, el estimado de la cópula fue de 4% abajo del valor verdadero y casi dentro del mismo rango que en las estimaciones para México. Para los hijos que nacieron en el quintil superior, 38% permanecieron ahí en la etapa adulta, para ambas NLSY. Sin embargo, las estimaciones de cópula con ρ_{Actual} son de 33% y 35%⁴⁶. Para hijos nacidos en el quintil superior, los valores de la NLSY son casi idénticos a nuestras estimaciones para México, aunque las estimaciones con la NLSY son más bajas.

La matriz de transición por deciles (Cuadro 8) muestra cuán grande es la brecha en la igualdad de oportunidades entre hijos nacidos en hogares pobres e hijos nacidos en hogares ricos. Al promediar a través de las dos cohortes mexicanas, estimamos que los hijos de los hogares del primer decil tienen una probabilidad casi 11 veces mayor de posicionarse en ese decil en su vida adulta en comparación con los hijos que provienen de hogares del decil superior. Además, los hijos en el decil superior tienen casi 10 veces más probabilidades de permanecer ahí en la edad adulta que los hijos que nacieron en los hogares del decil más bajo.⁴⁷ A partir de los datos de la NLSY para los Estados Unidos, las estimaciones resultan muy similares. Al promediar a través

de las dos NLSY, los hijos de hogares del decil más bajo tienen una probabilidad 10 veces mayor a la de los hijos nacidos en hogares del decil más alto de ubicarse en el decil más bajo durante su edad adulta (la probabilidad se incrementa a 12 veces por el método de cópula con ρ_{Actual}). Los hijos que provienen del decil más alto de la muestra NLSY tienen una probabilidad 14 veces mayor a los del decil inferior de posicionarse en su vida adulta en

Cuadro 7 Estimación de cópula de las matrices de transición por quintiles con errores estándar

Cohorte 1: 1984-2006 ($\rho_{Estimada} = 0.12$)

Padre	Hijo				
	Quintil 1	Quintil 2	Quintil 3	Quintil 4	Quintil 5
Quintil 1	34.6 (2.5)	26.2 (2.2)	17.5 (2.0)	13.6 (1.6)	7.9 (1.5)
Quintil 2	25.6 (2.5)	24.1 (2.0)	19.3 (2.2)	18.1 (2.1)	13.0 (1.8)
Quintil 3	19.0 (2.4)	20.9 (2.2)	21.0 (2.1)	20.9 (2.2)	18.0 (1.9)
Quintil 4	13.1 (1.8)	17.2 (2.2)	21.6 (2.4)	23.9 (2.3)	24.4 (2.0)
Quintil 5	7.4 (1.6)	11.8 (1.7)	20.4 (2.0)	23.7 (1.8)	36.9 (2.4)

Cohorte 2: 1989-2011 ($\rho_{Estimada} = 0.18$)

Padre	Hijo				
	Quintil 1	Quintil 2	Quintil 3	Quintil 4	Quintil 5
Quintil 1	37.9 (2.9)	23.8 (2.6)	17.6 (2.6)	12.6 (2.1)	8.0 (2.0)
Quintil 2	25.3 (2.4)	24.5 (2.7)	19.8 (2.6)	17.9 (2.3)	12.2 (2.1)
Quintil 3	17.7 (2.2)	21.9 (2.6)	21.4 (2.9)	22.1 (2.8)	17.1 (2.2)
Quintil 4	12.0 (2.1)	17.8 (2.2)	21.9 (2.5)	24.4 (2.3)	23.6 (2.3)
Quintil 5	7.0 (1.5)	11.9 (1.9)	19.5 (2.3)	22.8 (2.2)	38.5 (2.7)

Fuente: Estimación propia con base en los datos de la ENIGH 1984 y 1989 y la EMOVI 2006 y 2011.

46 Al usar datos de impuestos para hijos más o menos de la misma edad, tomados de la NLSY-1997, R. Chetty, *et al.*, «Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States», *NBER Working Paper 19843*, National Bureau of Economic Research, 2014 estiman que la probabilidad de que los hijos del quintil más bajo permanezcan en ese quintil es del 34%, en tanto que la probabilidad de los hijos del quintil más alto de quedarse en dicho quintil es del 37%. Sus muestras difieren de las nuestras en dos aspectos, a) ellos incluyen hogares con ingreso cero y b) no ajustan la equivalencia al usar el ingreso del hogar.

47 En la ENIGH84-EMOVI06, los hijos de hogares del decil más bajo tienen una probabilidad 15 veces mayor de posicionarse en el decil más bajo en la etapa adulta que los niños del decil más alto. De la ENIGH89-EMOVI11, estimamos que tienen 8 veces esta probabilidad. Al promediar las probabilidades de los niños en el decil más alto y más bajo a través de las dos cohortes, estimamos que los niños de hogares del decil más bajo tienen una probabilidad 11 veces mayor de quedarse en el decil más bajo. En cambio, la probabilidad de los hijos de permanecer en su vida adulta en el decil más alto es de 10 veces, tomada del promedio de las dos cohortes (11 veces en la cohorte 84-06 y 9 veces en la cohorte 89-11).

Cuadro 8

Estimación de cópula de las matrices de transición por deciles con errores estándar

Cohorte 1: 1984-2006 ($\rho_{Estimada} = 0.12$)

Padre Decil	Hijo									
	1	2	3	4	5	6	7	8	9	10
1	21.7 (3.3)	15.5 (2.7)	14.8 (2.7)	12.2 (2.7)	7.8 (2.2)	9.0 (2.4)	7.0 (2.2)	5.0 (1.6)	3.9 (1.4)	2.5 (1.2)
2	17.2 (3.0)	14.8 (2.5)	13.8 (2.7)	11.6 (2.3)	8.5 (2.3)	9.6 (2.2)	8.9 (2.2)	6.3 (1.9)	5.6 (1.8)	3.9 (1.6)
3	13.9 (2.8)	13.2 (2.9)	13.5 (2.5)	11.8 (2.5)	8.8 (2.3)	9.6 (2.4)	9.8 (2.6)	7.7 (2.2)	6.9 (2.2)	4.9 (1.6)
4	11.6 (2.6)	12.5 (3.2)	12.0 (2.8)	10.9 (2.4)	10.2 (2.4)	10.0 (2.6)	9.8 (2.2)	8.9 (2.7)	8.2 (2.4)	6.0 (2.1)
5	9.6 (2.5)	11.1 (2.3)	10.7 (2.4)	11.0 (2.6)	10.6 (2.6)	10.1 (2.2)	10.1 (2.1)	9.5 (2.3)	9.8 (2.6)	7.0 (2.1)
6	7.7 (2.2)	9.7 (2.3)	9.8 (2.4)	10.3 (2.3)	11.0 (2.5)	10.4 (2.5)	11.2 (2.8)	10.9 (2.5)	11.1 (2.5)	8.1 (2.3)
7	6.5 (1.9)	8.4 (2.3)	8.4 (2.4)	9.5 (2.6)	11.2 (2.8)	10.6 (2.4)	11.6 (2.5)	11.7 (2.9)	11.5 (2.3)	10.7 (2.3)
8	5.0 (1.8)	6.4 (2.0)	7.5 (2.1)	8.9 (2.2)	10.8 (2.6)	10.6 (2.7)	11.5 (2.7)	12.9 (2.6)	13.2 (2.5)	13.3 (2.6)
9	3.5 (1.6)	5.1 (1.8)	5.4 (1.7)	7.6 (2.2)	10.8 (2.3)	10.8 (2.4)	11.0 (2.5)	14.2 (2.4)	14.9 (2.9)	16.8 (2.6)
10	2.8 (1.5)	3.4 (1.6)	4.2 (1.6)	6.3 (1.9)	9.7 (2.2)	9.6 (2.1)	9.1 (2.2)	13.0 (2.7)	15.1 (2.8)	27.0 (2.9)

Cohorte 2: 1989-2011 ($\rho_{Estimada} = 0.18$)

Padre Decil	Hijo									
	1	2	3	4	5	6	7	8	9	10
1	28.0 (4.3)	14.1 (3.5)	12.1 (2.9)	12.1 (3.0)	8.3 (2.4)	7.8 (2.2)	5.3 (2.3)	5.3 (2.0)	3.8 (2.0)	3.0 (1.8)
2	19.3 (3.5)	14.3 (2.9)	11.8 (2.5)	11.7 (3.2)	9.6 (2.4)	9.4 (2.7)	7.5 (2.1)	7.1 (2.4)	5.6 (2.2)	3.6 (1.8)
3	14.0 (3.0)	13.3 (3.2)	12.4 (3.1)	12.5 (3.0)	9.6 (2.5)	9.6 (2.4)	8.7 (2.5)	8.0 (2.6)	6.9 (2.3)	4.6 (1.8)
4	10.9 (3.0)	12.5 (2.9)	12.5 (3.0)	11.5 (2.9)	10.2 (2.9)	10.2 (3.1)	10.1 (2.5)	9.0 (2.6)	7.5 (2.2)	5.3 (2.0)
5	8.3 (2.4)	10.9 (2.9)	11.7 (2.8)	11.1 (2.8)	10.5 (3.0)	10.6 (2.8)	10.5 (2.8)	10.6 (2.9)	8.6 (2.7)	7.0 (2.5)
6	6.4 (2.0)	9.8 (2.5)	10.7 (2.5)	10.3 (3.1)	10.7 (3.2)	11.1 (2.9)	11.8 (2.9)	11.3 (3.0)	10.0 (3.1)	8.6 (2.3)
7	4.9 (2.0)	8.3 (2.3)	9.3 (2.6)	9.5 (2.3)	10.4 (3.0)	11.7 (3.1)	12.2 (2.9)	11.9 (3.1)	10.7 (3.0)	10.7 (2.8)
8	3.4 (1.7)	7.3 (2.3)	8.6 (2.5)	8.3 (2.7)	10.5 (2.7)	11.3 (3.2)	11.9 (2.8)	12.7 (2.8)	12.9 (3.3)	12.9 (3.3)
9	2.8 (1.7)	5.4 (2.1)	6.9 (2.4)	7.3 (2.4)	10.7 (2.9)	10.7 (2.8)	11.3 (2.7)	12.5 (3.6)	14.7 (3.1)	17.2 (3.3)
10	1.8 (1.2)	4.0 (1.5)	3.9 (1.8)	5.5 (1.9)	9.1 (2.4)	8.6 (2.5)	10.5 (2.5)	11.3 (2.9)	18.8 (3.1)	26.3 (3.1)

Fuente: Estimación propia con base en los datos de la ENIGH 1984 y 1989 y la EMOVI 2006 y 2011.

el decil más rico (9 veces de acuerdo con la estimación de cópula con ρ_{Actual}).

Podemos también comparar la movilidad de los rangos entre México y los Estados Unidos al revisar la pendiente rango-rango, tal como lo proponen Chetty, Hendren, Kline, y Saez.⁴⁸ Éste es el coeficiente de una regresión de rangos de hijos sobre padres en sus respectivas distribuciones de ingreso. Nuestras estimaciones de la pendiente rango-rango para México son de 0.35 para la cohorte 1984-2006 y 0.38 para la cohorte 1989-2011. Esto es muy similar a las estimaciones de cópula de 0.39 para la muestra NLSY-1979 y 0.35 para la muestra NLSY-1997. Ambas estimaciones de cópula NLSY subestiman ligeramente las pendientes rango-rango de la muestra, la cual es de 0.42 en la muestra NLSY-1979 y 0.40 en la muestra NLSY-1997.⁴⁹

48 R. Chetty, et al., «Where Is...», op. cit.

49 Los resultados completos de la pendiente rango-rango y las estimaciones de cópula están en el Cuadro A1.4 en el apéndice 1. R. Chetty, et al., «Where Is...», op. cit. esti-

6. Conclusión

En este capítulo usamos distintas medidas de movilidad, tales como curvas de movilidad, IGE, correlación y matrices de transición, para mostrar que los paneles sintéticos pueden utilizarse para estimar con razonable precisión la movilidad intergeneracional de ingreso ante la ausencia de datos de panel. Al hacer esto, introducimos el uso de cópulas para mejorar la precisión de las estimaciones de movilidad por el método de panel sintético sobre los métodos que imponen fuertes supuestos de distribución y cuya invalidez se reconoce ampliamente. Validamos el uso de paneles sintéticos basados en cópulas sobre la movilidad intergeneracional e intrageneracional, y usamos la técnica de paneles sintéticos basados en cópulas

man que la pendiente rango-rango es de 0.341, usando para ello datos de impuestos, con las mismas diferencias que comentamos antes acerca de la inclusión de hogares con ingreso cero y no ajuste a la equivalencia del ingreso del hogar. Como contexto, ellos estiman un indicador de 0.180 para la pendiente rango-rango en Dinamarca, lo cual implica un grado de persistencia mucho más bajo en el rango de ingreso a través de las generaciones en Dinamarca de lo que ocurre en Estados Unidos o México.

para estimar la movilidad de ingreso en el contexto de México, donde no existen datos de panel.

Lo anterior nos permitió construir estimaciones para cada una de las medidas de movilidad de ingreso en México. Nuestras estimaciones nos facultaron para cuantificar las brechas en la igualdad de oportunidades en México, por ejemplo al comparar los prospectos de hijos nacidos en los hogares más ricos con aquéllos nacidos en los hogares más pobres. Estimamos que los hijos de los

hogares del decil más pobre tienen una probabilidad 11 veces mayor de quedarse ahí en su vida adulta que los que crecen en el decil más alto, de convertirse en adultos en el decil más bajo. También resulta 10 veces más probable que los hijos que provienen de los hogares del decil más rico se posicionen en su vida adulta en el decil más alto que los que nacieron en el decil más pobre. Finalmente, los resultados sugieren que la movilidad por rangos es muy similar en México y en Estados Unidos.

Referencias

- Antman, Francisca y David J. McKenzie, «Earnings Mobility and Measurement Error: A Pseudo-Panel Approach», *Economic Development and Cultural Change*, vol. 56, núm. 1, 2007, pp. 125–161.
- Banks, James, Richard Blundell y Agar Brugiavini, «Risk Pooling, Precautionary Saving and Consumption Growth», *The Review of Economic Studies*, vol. 68, núm. 4, 2001, pp. 757–779.
- Behrman, Jere R., Alejandro Gaviria, Miguel Székely, Nancy Birdsall y Sebastian Galiani, «Intergenerational Mobility in Latin America», *Economía*, vol. 2, núm. 1, 2001, pp. 1–44.
- Bø, Erlend E., Peter J. Lambert y Thor O. Thoresen, «Horizontal Inequity under a Dual Income Tax System: Principles and Measurement», *International Tax and Public Finance*, vol. 19, núm. 5, 2011, pp. 625–640.
- Bonhomme, Stephane y Jean-Marc Robin, «Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990–2000», *The Review of Economic Studies*, vol. 76, núm. 1, 2009, pp. 63–92.
- Cherubini, Umberto, Elisa Luciano y Walter Vecchiato, *Copula Methods in Finance*, West Sussex, John Wiley & Sons, 2004.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline y Emmanuel Saez, «Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States», *NBER Working Paper* 19843, National Bureau of Economic Research (NBER), 2014.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez y Nicholas Turner, «Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility», *nber Working Paper* 19844, National Bureau of Economic Research (nber), 2014.
- Cruces, Guillermo, Peter Lanjouw, Leonardo Lucchetti, Elizaveta Perova, Renos Vakis y Mariana Viollaz, «Intra-Generational Mobility and Repeated Cross-Sections: A Three-Country Validation Exercise», *Policy Research Working Paper* 5916, Banco Mundial, 2011.
- Dang, Hai-Anh y Peter Lanjouw, «Measuring Poverty Dynamics with Synthetic Panels Based on Cross-Sections», *Policy Research Working Paper* 6504, Banco Mundial, 2013.
- Dang, Hai-Anh, Peter Lanjouw, Jill Luoto y David McKenzie, «Using Repeated Cross-Sections to Explore Movements in and out of Poverty», *Policy Research Working Paper* 5550, Banco Mundial, 2011.
- Deaton, Angus, «Panel Data from Time Series of Cross-Sections», *Journal of Econometrics*, vol. 30, 1985, pp. 109–126.
- Demarta, Stefano y Alexander J. McNeil, «The T Copula and Related Copulas», *International Statistical Review*, vol. 73, núm. 1, 2007, pp. 111–129.
- Dunn, Christopher E., «The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil», *The BE Journal of Economic Analysis & Policy*, vol. 7, núm. 2, 2007, pp. 1–42.
- Elbers, Chris, Jean O. Lanjouw y Peter Lanjouw, «Micro-Level Estimation of Poverty and Inequality», *Econometrica*, vol. 71, núm. 1, 2003, pp. 355–364.
- Emran, M. Shahe y Forhad Shilpi, «Intergenerational Occupational Mobility in Rural Economy», *The Journal of Human Resources*, vol. 46, núm. 2, 2011, pp. 427–458.
- Emran, M. Shahe y Forhad Shilpi, «Gender, Geography and Generations Intergenerational Educational Mobility in Post-Reform India», *Policy Research Working Paper* 6055, Banco Mundial, 2012.
- Emran, M. Shahe y Yan Sun, «Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 1988–2002», *Working Paper*, 2011. Disponible en: <http://ssrn.com/abstract=1891572>
- Fields, Gary S. y Mariana Viollaz, «Can the Limitations of Panel Datasets Be Overcome by Using Pseudo-Panels to Estimate Income Mobility», *Working Paper*, 2013.
- Foster, James E. y Jonathan Rothbaum, «The Mobility Curve: Measuring the Impact of Mobility on Welfare», *Working Paper*, 2014. Disponible en: http://home.gwu.edu/~jrothbau/Rothbaum_Jonathan_MobilityCurves.pdf
- Guimaraes Ferreira, Sergio y Fernando A. Veloso, «Intergenerational Mobility of Wages in Brazil», *Brazilian Review of Econometrics*, vol. 26, núm. 2, 2006, pp.181–211.
- Hertz, Thomas, «Education, Inequality and Economic Mobility in South Africa», tesis doctoral, University of Massachusetts, 2001.
- Hertz, Thomas, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith y Alina Verashchagina, «The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends», *The BE Journal of Economic Analysis & Policy*, vol. 7, núm. 2, 2007, pp. 1–48.

Hnatkovska, Viktoria, Amartya Lahiri y Sourabh B. Paul, «Breaking the Caste Barrier Intergenerational Mobility in India», *Journal of Human Resources*, vol. 48, núm. 2, 2013, pp. 435–473.

Jääntti, Markus, Eva Sierminska y Philippe Van Kerm, «Modelling the Joint Distribution of Income and Wealth», *Working Paper*, International Association for Research in Income and Wealth (IARIW), 2012.

Jaworski, Piotr, Fabrizio Durante, Wolfgang Härdle y Tomasz Rychlik (eds.), *Copula Theory and Its Applications*, Nueva York, Springer, 2010.

Madrian, Brigitte C. y Lars John Lefgren, «An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents», *Journal of Economic and Social Measurement*, vol. 26, 2000, pp. 31–62.

Mazumder, Bhashkar, «Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data», *Review of Economics and Statistics*, vol. 87, núm. 2, 2005, pp. 235–255.

Nimubona, Alain-Désiré y Désiré Vencatachellum, «Intergenerational Education Mobility of Black and White South Africans», *Journal of Population Economics*, vol. 20, núm. 1, 2007, pp. 149–182.

Núñez, Javier I. y Leslie Miranda, «Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile», *The BE Journal of Economic Analysis & Policy*, vol. 10, núm. 1, 2010, pp. 1-15.

Quheng, Deng, Björn Gustafsson y Li Shi, «Intergenerational Income Persistence in Urban China», *IZA Discussion Paper 6907*, Institute for the Study of Labor (IZA), 2012.

Torche, Florencia, «Unequal But Fluid: Social Mobility in Chile in Comparative Perspective», *American Sociological Review*, vol. 70, núm. 3, 2005, pp. 422–450.

Trivedi, Pravin K. y David M. Zimmer, «Copula Modeling: An Introduction for Practitioners», *Foundations and Trends in Econometrics*, vol. 1, núm. 1, 2005, pp. 1-111.

US Census Bureau, «Estimating ASEC Variances with Replicate Weights», 2013. Disponible en: thedataweb.rm.census.gov/ftp/cps_ftp.html

Vélez-Grajales, Roberto, Raymundo Campos-Vázquez y Juan Enrique Huerta-Wong, *Informe Movilidad Social en México 2013. Imagina tu futuro*, México, Centro de Estudios Espinosa Yglesias (CEEY), 2013.

Vinh, Andrea, William E. Griffiths y Duangkamon Chotikapanich, «Bivariate Income Distributions for Assessing Inequality and Poverty under Dependent Samples», *Economic Modelling*, vol. 27, núm. 6, 2010, pp. 1473–1483.

Zimmer, David M. y H. Youn Kim, «The Dependence Structure of Income Distribution», *Applied Economics*, vol. 44, núm. 27, 2012, pp. 3573–3583.

Apéndice 1 Estimaciones de panel sintético de otras medidas de movilidad

Cuadro a1.1

Comparación de la elasticidad intergeneracional del ingreso verdadera con las estimaciones de panel sintético

	Verdadera	DLLM			Cópula		
		ρ_L	ρ_{Actual}	ρ_H	ρ_L	ρ_{Actual}	ρ_H
PSID-CNEF (2005-2007)	0.76	0.63 0.77	0.78 0.87	0.88 0.94	0.63 0.71	0.78 0.86	0.88 0.96
CPS ASEC (2005-2006)	0.67	0.58 0.67	0.68 0.75	0.78 0.82	0.58 0.63	0.68 0.71	0.78 0.79
NLSY-1979	0.44	0.009 0.24	0.24 0.35	0.39 0.47	0.009 0.32	0.24 0.43	0.39 0.54
NLSY-1997	0.33	0.009 0.21	0.24 0.31	0.39 0.41	0.009 0.23	0.24 0.33	0.39 0.44
ENIGH-1984 / EMOVI-2006		0.05 0.26	0.12 0.30	0.30 0.42	0.05 0.32	0.12 0.35	0.30 0.45
ENIGH-1989 / EMOVI-2011		0.05 0.23	0.18 0.27	0.30 0.37	0.05 0.28	0.18 0.33	0.30 0.40

Cuadro a1.2

Comparación de la correlación del logaritmo del ingreso verdadera con las estimaciones de panel sintético

	Verdadera	DLLM			Cópula		
		ρ_L	ρ_{Actual}	ρ_H	ρ_L	ρ_{Actual}	ρ_H
PSID-CNEF (2005-2007)	0.70	0.63 0.74	0.78 0.84	0.88 0.91	0.63 0.65	0.78 0.78	0.88 0.88
CPS ASEC (2005-2006)	0.67	0.58 0.67	0.68 0.75	0.78 0.83	0.58 0.65	0.68 0.73	0.78 0.81
NLSY-1979	0.36	0.009 0.20	0.24 0.31	0.39 0.42	0.009 0.25	0.24 0.35	0.39 0.46
NLSY-1997	0.31	0.009 0.20	0.24 0.32	0.39 0.44	0.009 0.20	0.24 0.32	0.39 0.43
ENIGH-1984 / EMOVI-2006		0.05 0.31	0.12 0.34	0.30 0.44	0.05 0.27	0.12 0.39	0.30 0.49
ENIGH-1989 / EMOVI-2011		0.05 0.27	0.18 0.30	0.30 0.41	0.05 0.36	0.18 0.42	0.30 0.48

Notas: Estos cuadros comparan la correlación del logaritmo del ingreso verdadera con la estimación de los paneles sintéticos basados en DLLM y cópula Gaussiana usando la correlación conocida del error mco (ρ_{Actual}) y un límite plausible, inferior y superior, de las correlaciones posibles (ρ_L y ρ_H). Para las estimaciones de la movilidad intergeneracional en México utilizando la ENIGH84-EMOVI06 (nacidos en 1966-1976) y ENIGH89-EMOVI11 (nacidos en 1971-1981), la verdadera ρ no es conocida, por lo cual se especifica una correlación estimada del error: $\rho_{Estimada}$.

Fuente: Estimación propia con base en los datos de la ENIGH 1984 y 1989 y la EMOVI 2006 y 2011.

Cuadro a1.3 Comparación de las matrices de transición por quintiles

PSID-CNEF (2005-2007)

Verdadera

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	65.7	22.5	7.4	3.7	0.4
ii	23.0	47.9	22.1	6.0	1.3
iii	5.8	21.6	45.5	20.0	7.2
iv	3.1	5.7	21.2	52.0	17.9
v	2.4	2.3	3.7	18.1	73.5

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	60.1, 68.9, 76.9	25.1, 23.7, 20.6	10.7, 6.3, 2.3	3.5, 0.9, 0.1	0.5, 0.0, 0.0
ii	25.3, 24.0, 20.4	32.9, 39.6, 49.8	24.5, 26.1, 25.5	13.3, 9.3, 4.2	3.9, 1.0, 0.1
iii	10.6, 6.0, 2.4	24.9, 26.3, 25.1	28.5, 35.1, 44.6	24.8, 26.0, 25.5	11.2, 6.6, 2.5
iv	3.3, 0.9, 0.1	13.5, 9.4, 4.5	25.1, 26.0, 25.1	32.4, 39.4, 49.0	25.8, 24.2, 21.3
v	0.5, 0.1, 0.0	3.6, 1.0, 0.1	11.2, 6.4, 2.5	26.0, 24.4, 21.2	58.8, 68.3, 76.3

cópula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	55.5, 66.1, 74.5	25.4, 24.0, 21.5	12.4, 7.6, 3.4	5.4, 2.0, 0.4	1.2, 0.2, 0.0
ii	24.9, 22.8, 20.4	31.2, 37.8, 46.4	24.5, 26.0, 26.4	14.6, 11.5, 6.4	4.7, 1.9, 0.4
iii	12.3, 8.2, 4.2	23.8, 24.9, 24.8	27.8, 33.3, 40.7	24.3, 25.4, 26.4	11.8, 8.1, 3.9
iv	5.7, 2.5, 0.7	14.5, 11.0, 6.8	23.9, 25.4, 25.7	30.6, 37.1, 45.0	25.3, 24.0, 21.9
v	1.5, 0.3, 0.0	5.0, 2.2, 0.5	11.5, 7.8, 4.0	25.0, 24.0, 21.7	57.1, 65.8, 73.9

CPS ASEC (2005-2006)

Verdadera

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	64.6	21.3	8.0	3.7	2.4
ii	21.6	43.0	20.7	9.6	5.1
iii	7.8	21.0	40.8	20.6	9.9
iv	3.7	9.0	21.2	44.7	21.5
v	2.3	5.8	9.3	21.3	61.2

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	54.5, 60.3, 67.1	26.0, 25.5, 24.4	13.0, 10.5, 7.2	5.4, 3.3, 1.3	1.2, 0.4, 0.1
ii	26.0, 25.5, 24.4	29.6, 33.0, 38.2	23.7, 24.8, 25.9	15.3, 13.4, 10.2	5.4, 3.3, 1.3
iii	12.9, 10.5, 7.2	23.7, 24.8, 25.9	26.7, 29.4, 34.0	23.7, 24.8, 25.8	13.0, 10.5, 7.2
iv	5.4, 3.3, 1.3	15.3, 13.5, 10.2	23.7, 24.8, 25.8	29.7, 33.0, 38.2	26.0, 25.4, 24.5
v	1.2, 0.4, 0.1	5.5, 3.3, 1.4	13.0, 10.5, 7.2	26.0, 25.5, 24.5	54.4, 60.3, 66.9

cópula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	54.5, 61.3, 68.3	25.3, 24.8, 23.4	12.3, 9.9, 6.7	5.5, 3.4, 1.5	1.4, 0.6, 0.1
ii	26.0, 25.4, 23.8	30.2, 33.5, 39.1	23.5, 24.5, 25.3	14.7, 13.1, 10.3	5.6, 3.5, 1.6
iii	12.4, 9.8, 6.6	24.1, 25.0, 25.7	27.0, 29.9, 34.4	23.6, 24.6, 25.6	13.0, 10.7, 7.7
iv	5.0, 3.1, 1.3	15.0, 13.2, 10.3	24.1, 25.1, 26.0	30.1, 33.1, 38.0	25.9, 25.5, 24.5
v	1.2, 0.5, 0.1	5.5, 3.4, 1.5	13.1, 10.7, 7.7	26.1, 25.7, 24.6	54.1, 59.7, 66.1

- Valores verdaderos que se encuentran dentro del rango [ρ_L, ρ_H].
- Valores verdaderos que se encuentran dentro del intervalo de confianza al 95% de la estimación de ρ_{Actual} .
- Valores verdaderos que se encuentran dentro con un 95% de confianza de ρ_{Actual} y del rango [ρ_L, ρ_H].

NLSY-1979

Verdadera

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	43.2	24.8	16.0	10.6	5.3
ii	22.8	24.2	22.9	16.8	13.3
iii	13.8	21.6	24.2	21.6	18.9
iv	13.0	16.7	18.7	26.4	25.2
v	7.1	12.9	18.2	24.4	37.5

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	29.8, 34.6, 39.8	22.6, 23.6, 24.7	19.2, 18.4, 17.4	16.0, 14.3, 12.0	12.3, 9.1, 6.0
ii	22.3, 23.6, 24.7	21.2, 22.3, 23.8	20.4, 20.5, 21.3	19.2, 18.9, 18.0	16.9, 14.8, 12.1
iii	18.9, 18.3, 17.4	19.9, 20.6, 21.4	20.4, 21.3, 21.8	20.4, 20.6, 21.4	20.4, 19.2, 18.0
iv	16.1, 14.2, 11.8	19.0, 18.5, 18.0	20.1, 20.7, 21.3	21.6, 22.5, 23.9	23.1, 24.1, 25.1
v	12.8, 9.4, 6.3	17.3, 15.0, 12.1	19.8, 19.1, 18.2	22.9, 23.7, 24.7	27.3, 32.8, 38.8

cópula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	33.2, 38.5, 44.0	22.3, 23.1, 23.6	19.0, 17.6, 16.6	14.9, 12.9, 10.5	10.6, 7.8, 5.2
ii	23.6, 24.3, 25.3	21.7, 22.9, 24.4	20.3, 21.0, 21.2	18.5, 18.3, 17.7	15.9, 13.5, 11.5
iii	18.6, 17.8, 16.3	20.5, 20.8, 21.5	20.4, 20.9, 22.1	20.9, 21.4, 22.2	19.6, 19.1, 17.9
iv	14.4, 12.4, 9.9	19.0, 18.6, 18.4	20.4, 21.1, 21.7	22.4, 23.1, 24.6	23.8, 24.8, 25.4
v	10.1, 6.9, 4.4	16.6, 14.6, 12.1	20.0, 19.4, 18.4	23.3, 24.3, 25.0	30.2, 34.9, 40.1

NLSY-1997

verdadera

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	39.9	22.8	17.3	11.4	8.5
ii	21.7	25.7	22.1	19.3	11.3
iii	17.3	22.5	21.5	19.7	19.1
iv	11.5	16.5	24.3	23.9	23.6
v	9.5	12.6	14.8	25.7	37.6

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	30.0, 35.0, 40.9	22.5, 24.0, 24.9	18.9, 18.4, 17.0	16.1, 13.6, 11.5	12.4, 8.9, 5.7
ii	22.6, 23.9, 24.9	21.4, 22.4, 24.3	20.3, 20.6, 21.4	18.9, 18.6, 17.7	16.8, 14.6, 11.7
iii	18.6, 18.1, 17.2	20.1, 20.6, 21.4	20.4, 21.0, 22.3	20.6, 21.1, 21.3	20.3, 19.2, 17.8
iv	15.9, 13.9, 11.2	18.7, 18.2, 17.6	20.6, 20.9, 21.6	21.7, 22.6, 24.4	23.1, 24.4, 25.2
v	12.8, 9.1, 5.8	17.2, 14.8, 11.9	19.9, 19.1, 17.6	22.7, 24.1, 25.2	27.3, 32.8, 39.5

cópula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Padres (quintiles)	Hijos (quintiles)				
	i	ii	iii	iv	v
i	30.1, 35.5, 41.4	22.1, 23.3, 24.1	18.9, 18.2, 16.8	16.1, 13.8, 11.7	12.7, 9.1, 6.0
ii	23.4, 24.2, 25.2	21.7, 22.7, 24.2	20.2, 20.5, 21.3	18.6, 18.4, 17.6	16.0, 14.2, 11.8
iii	18.8, 18.0, 16.8	20.1, 20.7, 21.6	20.6, 21.4, 22.3	20.4, 20.9, 21.6	20.1, 19.0, 17.7
iv	15.6, 13.4, 11.0	18.9, 18.4, 18.1	20.6, 21.1, 21.8	21.7, 22.9, 24.2	23.2, 24.2, 25.0
v	12.1, 8.8, 5.5	17.1, 14.8, 12.1	19.6, 18.8, 17.9	23.2, 24.0, 24.9	27.9, 33.4, 39.5

Notas: Este cuadro compara la matriz de transiciones por quintiles verdadera con las estimaciones de los paneles sintéticos basados en DLLM y en la cópula Gaussiana usando la correlación conocida de los errores MCO (ρ_{Actual}) y un límite plausible inferior y superior de las posibles correlaciones (ρ_L y ρ_H). Las estimaciones de panel sintético para cada transición de quintil a quintil son $\rho_L, \rho_{Actual}, \rho_H$. Los errores estándar ρ_{Actual} fueron generados por el método *bootstrap* con 100 réplicas.
Fuente: Estimación propia.

Cuadro a1.4
Pendiente rango-rango

PSID-CNEF				CPS ASEC				ENIGH-1984/EMOVI-2006				ENIGH-1989/EMOVI-2011			
ρ	verd.	DLLM	cópula	ρ	verd.	DLLM	cópula	ρ	verd.	DLLM	cópula	ρ	verd.	DLLM	cópula
0.63	0.820	0.737	0.678	0.58	0.720	0.655	0.644	0.05	N/A	0.237	0.321	0.05	N/A	0.233	0.316
0.68		0.770	0.712	0.63		0.694	0.682	0.10		0.265	0.343	0.10		0.260	0.344
0.73		0.805	0.745	0.68		0.734	0.721	0.12		0.272	0.354	0.12		0.272	0.349
0.78		0.841	0.783	0.73		0.774	0.761	0.15		0.295	0.372	0.15		0.287	0.363
0.83		0.875	0.823	0.78		0.815	0.802	0.18		0.313	0.386	0.18		0.304	0.376
0.88		0.912	0.865					0.20		0.326	0.394	0.20		0.318	0.394
NLSY-1979				NLSY-1997				0.25		0.348	0.418	0.25		0.349	0.416
ρ	verd.	DLLM	cópula	ρ	verd.	DLLM	cópula	0.30		0.383	0.446	0.30		0.383	0.440
0.09	0.418	0.218	0.288	0.09	0.403	0.222	0.243								
0.14		0.248	0.321	0.14		0.258	0.279								
0.19		0.282	0.352	0.19		0.294	0.314								
0.24		0.316	0.386	0.24		0.327	0.350								
0.29		0.349	0.416	0.29		0.363	0.382								
0.34		0.382	0.449	0.34		0.399	0.418								
0.39		0.416	0.484	0.39		0.436	0.454								

Valores verdaderos que se encuentran dentro del intervalo al 95% de confianza.

No hay diferencia estadísticamente significativa entre el valor verdadero y el estimado.

El valor verdadero se encuentra dentro de ρ_{Actual} al 95% de confianza y no hay diferencia estadísticamente significativa con el estimado.

Notas: La pendiente rango-rango es el coeficiente de la regresión del rango del ingreso del hijo sobre el rango del ingreso del padre. Este cuadro compara el valor verdadero y la estimación de panel sintético de la pendiente rango-rango para cada conjunto de datos para distintos valores ρ incluyendo la correlación verdadera del error cuando es conocida y la correlación estimada para la ENIGH-1984/EMOVI-2006 (nacidos en 1966-1976) y la ENIGH-1989/EMOVI-2011 (nacidos en 1971-1981).

Fuente: Estimación propia.

