



Emparejamiento selectivo: medición y tendencias en América Latina y el Caribe

Iván Albina

Tesis de Maestría

Maestría en Economía

Universidad Nacional de La Plata

Directores de tesis: Lucila Berniell y Leonardo Gasparini

Fecha de defensa: 03/04/2025

Códigos JEL: I24, J11, J12

Emparejamiento selectivo: medición y tendencias en América Latina y el Caribe*

Iván Albina*

Resumen

Este trabajo documenta el grado y la evolución del emparejamiento selectivo (ES) educativo en 21 países de América Latina y el Caribe para cohortes nacidas entre 1915 y 1985, utilizando microdatos censales. Para estudiar tendencias en un contexto de expansión educativa y cambios en las distribuciones marginales por género, se evalúan indicadores de ES existentes y se propone el Indicador Cuantílico Educativo Local (ICEL), definido como la razón entre la probabilidad observada de emparejamiento dentro de un mismo cuantil educativo y la probabilidad esperada bajo independencia estadística. Se presentan dos extensiones: un indicador global que incorpora en la medición la selección hacia la formación de pareja y un indicador geográfico que integra heterogeneidad territorial en capital educativo y patrones de emparejamiento en mercados matrimoniales locales. De acuerdo con los datos de los últimos censos disponibles de cada país, la probabilidad de emparejarse con una persona del mismo quintil educativo es, en promedio, 2.3 veces la que surgiría bajo emparejamiento aleatorio, y presenta una mayor intensidad en los extremos de la distribución (Q1 en torno a 2.5; Q5 en torno a 3). Entre el primer y el último registro por país, el ICEL disminuye en 13 de 21 casos. Simulaciones para México muestran que restricciones etarias y geográficas explican una fracción sustantiva del emparejamiento observado y que, en escenarios contrafactuales, la reducción del Gini bajo aleatorización completa (6%) se atenúa a 3% al incluir restricciones etarias, geográficas y maritales.

Palabras clave: emparejamiento selectivo, mercados matrimoniales, educación, IPUMS, censos, América Latina y el Caribe

Códigos JEL: I24, J11, J12

* Este trabajo constituye mi Tesis de Maestría en Economía de la Universidad Nacional de La Plata. Agradezco especialmente a Leonardo Gasparini y a Lucila Berniell por su guía y apoyo durante todo el proceso. También agradezco a Mariana Marchionni, Joaquín Coleff, Luis Laguinge y Facundo Crosta por sus valiosos comentarios en el Seminario de Avance de Tesis realizado el 1 de agosto de 2024. Los errores u omisiones son de mi exclusiva responsabilidad.

* ORCID: <https://orcid.org/0009-0000-2732-4433>. Centro de Estudios Distributivos, Laborales y Sociales del Instituto de Investigaciones Económicas de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata; La Plata, Argentina. Mail: ivan_albina@hotmail.com

1. Introducción

Durante la segunda mitad del siglo XX, América Latina y el Caribe (ALC) experimentó transformaciones demográficas, educativas y urbanas de gran magnitud. La expansión educativa —en particular entre las mujeres—, junto con la urbanización y la diversificación de las formas de unión, coincidió con cambios en la composición de los potenciales cónyuges y en los contextos sociales y espaciales en los que se forman las parejas (Esteve et al., 2012; García & Rojas, 2004; Quilodrán, 2001). En este contexto, el emparejamiento selectivo (ES)—la tendencia a formar pareja con individuos similares en determinadas características en una magnitud mayor que la que surgiría bajo emparejamiento aleatorio— emerge como un fenómeno central para caracterizar la asignación de capital humano entre hogares y, por esa vía, su vínculo con la desigualdad y la transmisión intergeneracional.¹²

El emparejamiento selectivo educativo es relevante porque puede amplificar la desigualdad entre hogares de al menos dos maneras. Por un lado, cuando individuos con niveles educativos más altos (y, por extensión, con mayor capacidad potencial de generación de ingresos) tienden a emparejarse entre sí, los recursos se concentran en ciertos hogares y se expanden las brechas respecto de aquellos donde ambos miembros tienen menor capacidad de generación de ingresos. Por otro lado, a través de la transmisión intergeneracional de capital humano y atributos económicos (y otros rasgos correlacionados), el ES puede contribuir a perpetuar niveles elevados de desigualdad y limitar la movilidad social en generaciones futuras (Torche, 2010, 2014). En contextos de alta desigualdad inicial como los prevalecientes en América Latina y el Caribe, estas dinámicas adquieren particular relevancia.

A pesar de su potencial relevancia para la dinámica de la desigualdad, la evidencia disponible no permite establecer un consenso sobre su evolución reciente. Este resultado se explica, en parte, por el hecho de que la medición del mismo entraña dificultades de distinta naturaleza y aún no existe un acuerdo sobre la métrica más adecuada para evaluarlo (Bratsberg et al., 2023; Chiappori et al., 2021, 2025). El desafío es metodológico: las variaciones observadas en los patrones de formación de parejas pueden originarse en fuerzas conceptualmente diferentes. Por un lado, (i) cambios en la composición educativa de la población —variaciones en las distribuciones marginales de educación de hombres y mujeres— modifican el conjunto de combinaciones posibles en el mercado matrimonial y pueden inducir cambios en el emparejamiento aun cuando el proceso subyacente de búsqueda y selección permanezca inalterado. Por otro lado, (ii) los cambios pueden reflejar transformaciones en la “tecnología de emparejamiento”, es decir, en los mecanismos e instituciones que estructuran los encuentros y la formación de parejas —incluyendo normas sociales, preferencias, segmentación de redes y barreras que facilitan o restringen vínculos entre grupos— (Becker, 1974). Distinguir empíricamente entre ambas fuentes de

¹ En particular, el *emparejamiento selectivo positivo* se refiere a que la similitud entre cónyuges en un rasgo (p. ej., educación) es mayor que la esperada bajo emparejamiento aleatorio. En la práctica, gran parte de la evidencia empírica estudia ES a partir de la educación por su medición relativamente estable y comparable en el tiempo, aunque el interés suele estar en su vínculo con el potencial de ingresos y la desigualdad del ingreso del hogar.

² En este trabajo, el universo de parejas se restringe a aquellas formadas por un hombre y una mujer. Si bien reconocemos a las parejas homosexuales, la fuente de datos utilizada no permite identificar este tipo de parejas. Igualmente, basándonos en la EPH-C de Argentina, se puede observar que el porcentaje de hogares conformados por parejas homosexuales era de 0.2% en 2003-S2, y asciende a 1% en 2023-S2, con lo cual el tamaño de la muestra sería bastante acotado para estudiar este fenómeno.

variación resulta central para evaluar las implicancias del ES sobre la desigualdad y la movilidad social, especialmente en comparaciones temporales e internacionales.

Más allá de los desafíos de medición, comprender el rol del emparejamiento en la desigualdad requiere considerar dimensiones que suelen pasar desapercibidas en el estudio del fenómeno. Primero, el ES puede diferir a lo largo de la distribución educativa: no es equivalente que la selección se concentre en los extremos a que lo haga en niveles intermedios, ya que sus implicancias distributivas pueden ser distintas. Segundo, la formación de pareja no es un proceso aleatorio en la población: quienes se emparejan pueden presentar características sistemáticamente diferentes de quienes no lo hacen, por lo que ignorar esta selección puede sesgar la medición de su evolución, en particular en contextos donde disminuye la proporción de personas en pareja (*unmating*), como es el caso de la región en los últimos cincuenta años.³ Finalmente, el nivel educativo y los patrones de emparejamiento pueden exhibir heterogeneidad espacial, de modo que su intensidad y sus consecuencias varíen según el contexto regional. En América Latina y el Caribe, donde las desigualdades territoriales son significativas, la movilidad interna es relativamente baja y las distancias geográficas pueden constituir fricciones relevantes en mercados matrimoniales segmentados (De La Mata et al., 2022), esta dimensión adquiere una importancia adicional.

Este trabajo contribuye en cuatro frentes. Primero, evalúa críticamente las métricas más utilizadas en la literatura para medir emparejamiento selectivo educativo y, para el estudio de tendencias, enfatiza el uso de enfoques basados en posiciones relativas. En ese marco, propone el Indicador Cuantílico Educativo Local (ICEL), definido como el cociente entre (i) la probabilidad observada de emparejarse dentro de un mismo cuantil educativo y (ii) la probabilidad esperada bajo independencia estadística. Su construcción busca ser robusta a cambios en las distribuciones marginales de educación y en las brechas de género, y permite describir heterogeneidad a lo largo de la distribución educativa. Segundo, introduce dos extensiones del ICEL: un indicador global que incorpora la selección asociada a la decisión de estar en pareja y un indicador geográfico que integra la heterogeneidad territorial en el capital educativo y el emparejamiento dentro de los mercados matrimoniales locales. Tercero, utilizando microdatos censales armonizados de IPUMS-International (IPUMS, 2020), documenta el nivel y la evolución del ES educativo en 21 países de América Latina y el Caribe para cohortes nacidas entre 1915 y 1985, a partir de seis rondas censales (1960–2010)⁴. Cuarto, cuantifica—para el caso de México—la relevancia empírica de restricciones demográficas, etarias y geográficas en el emparejamiento observado y evalúa implicancias distributivas mediante ejercicios contrafactuales de emparejamiento aleatorio bajo distintos conjuntos de restricciones.

Los resultados se organizan en tres bloques. Primero, la comparación entre indicadores muestra que varias métricas empleadas en la literatura no logran aislar adecuadamente los cambios en patrones de asignación de los cambios inducidos por variaciones en las distribuciones marginales educativas (y, en algunos casos, por cambios en brechas de género), lo que limita su uso para comparaciones intertemporales. En este marco, los enfoques basados en posiciones relativas —junto con el ICEL— presentan mejores propiedades para el análisis de tendencias y para la comparación entre países.

³ Mientras que, en la década del 1960, cerca del 65% de hombres y mujeres de entre 25 y 45 años se encontraba en pareja, este porcentaje ascendía a un 53% en promedio, en la década de 2010.

⁴ Para México se incluye además 2020.

Segundo, sobre la base de estos indicadores, se documenta ES educativo positivo en los 21 países y una heterogeneidad marcada por posición en la distribución: en la última ronda censal disponible, la probabilidad de emparejarse con alguien del mismo quintil es, en promedio, 2.3 veces la que resultaría bajo emparejamiento aleatorio, con mayor intensidad en los extremos (Q1 en torno a 2.5; Q5 en torno a 3). En términos de tendencias, entre el primer y el último registro por país, el ICEL disminuye en 13 de 21 casos. Tercero, al incorporar las extensiones del indicador, se muestra que la selección hacia la formación de pareja y la heterogeneidad espacial aportan información adicional para interpretar niveles y cambios del emparejamiento y sus implicancias distributivas. En México, las simulaciones sugieren que restricciones etarias y geográficas explican una fracción sustantiva del emparejamiento observado y que, en ejercicios contrafactuales, la reducción del Gini bajo aleatorización completa (aproximadamente 6%) se atenúa a aproximadamente 3% al imponer restricciones consistentes con mercados matrimoniales locales.

Este trabajo se relaciona con un cuerpo de literatura empírica que ha medido el ES desde la posguerra hasta la actualidad en países desarrollados, principalmente en Estados Unidos, y ha documentado tendencias divergentes según la métrica empleada. La literatura inicial cuantificó el ES sin incorporar adecuadamente el rol de cambios en las distribuciones marginales educativas (Breen & Salazar, 2011; Fernandez et al., 2005; Ganguli et al., 2014; Greenwood et al., 2014; Kremer, 1997). Un segundo grupo de trabajos avanzó al introducir normalizaciones y contrafacticos de emparejamiento aleatorio, aunque con dificultades para controlar completamente cambios en brechas de género y composición interna de los grupos educativos (Eika et al., 2019; Schwartz & Mare, 2005; Shen, 2019). Más recientemente, surgió una tercera generación de métricas con mejores propiedades para comparaciones intertemporales (Bratsberg et al., 2023; Chiappori et al., 2021; Gihleb & Lang, 2020; Handy, 2021). Este documento contribuye a esa agenda evaluando de manera sistemática el desempeño de métricas existentes, proponiendo un indicador cuantílico robusto para el estudio de tendencias y aportando evidencia para un conjunto amplio de países de América Latina y el Caribe en un período extenso. El trabajo también contribuye a la literatura que estudia el vínculo entre emparejamiento selectivo y desigualdad a través de ejercicios contrafactuales de formación aleatoria de hogares. Esta dimensión ha recibido atención en trabajos para Estados Unidos (Ciscato & Weber, 2020; Eika et al., 2019; Greenwood et al., 2014; Harmenberg, 2014) y para países de América Latina (Funes Leal, 2015; González & Dip, 2022; Pereira et al., 2020, entre otros). Este trabajo amplía la evidencia para el caso mexicano y pone especial énfasis en cómo las restricciones demográficas y geográficas incluidas en los contrafacticos afectan la magnitud del efecto sobre la desigualdad.

El resto del documento se organiza de la siguiente manera. La Sección 2 describe datos y estadísticas descriptivas. La Sección 3 presenta desafíos de medición y metodología (ICEL y extensiones). La Sección 4 muestra resultados comparados para la región. La Sección 5 cuantifica el rol de restricciones demográficas, maritales y geográficas en México. La Sección 6 evalúa implicancias distributivas bajo contrafactuales con distintas restricciones. La Sección 7 concluye.

2. Datos y estadísticas descriptivas

2.1. Fuentes de datos

Para estudiar el nivel y la evolución del emparejamiento selectivo en América Latina y el Caribe, se utilizan microdatos censales armonizados por IPUMS *International* (IPUMS, 2020) para 21 países de la región⁵⁶, con relevamientos que van desde la década de 1960 hasta la de 2010 (los censos específicos por país se detallan en la Tabla A1 del Apéndice). IPUMS-I provee información demográfica y socioeconómica a nivel individual y ofrece variables educativas armonizadas comparables internacionalmente, incluyendo el máximo nivel educativo alcanzado y los años de escolaridad. Los censos, a su vez, incorporan identificadores de hogar que permiten vincular cónyuges dentro de una misma unidad residencial; el procedimiento de emparejamiento se describe en el Apéndice I.

Adicionalmente, para el análisis espacial se emplean identificadores geográficos armonizados a nivel subnacional (primer y segundo nivel administrativo), que permiten comparaciones consistentes en el tiempo dentro de cada país.⁷ En conjunto, los censos resultan especialmente adecuados para el estudio del emparejamiento selectivo educativo por tres razones: su amplia cobertura temporal, su representatividad nacional con desagregación espacial, y el gran tamaño muestral, que posibilita estimar métricas del fenómeno incluso para unidades geográficas pequeñas, interpretables como aproximaciones a mercados matrimoniales locales.

2.2. Estadísticas descriptivas

El análisis principal se centra en parejas co-residentes en América Latina y el Caribe, definidas como parejas que conviven en el mismo hogar, en las que al menos uno de los miembros tiene entre 25 y 45 años. La Tabla 1 presenta estadísticas descriptivas para hombres y mujeres en ese rango etario, distinguiendo entre quienes se encuentran en pareja y quienes permanecen solteros o separados (la Tabla A2 del Apéndice reproduce estas estadísticas para cada década).

La Tabla 1 resalta tres hechos estilizados. En primer lugar, evidencia la marcada expansión educativa en la región a lo largo del siglo XX, reflejada en un aumento sostenido de los años de escolaridad promedio, impulsado por la caída en la proporción de individuos con primaria incompleta y el incremento de los niveles educativos más altos.⁸ En segundo lugar, muestra que, si bien la expansión educativa fue generalizada, el avance fue relativamente mayor entre las mujeres, lo que redujo la brecha educativa por género e

⁵ Los países considerados en el análisis son: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Haití, Honduras, Jamaica, México, Nicaragua, Panamá, Paraguay, Perú, Puerto Rico, República Dominicana, Uruguay y Venezuela. Si bien se dispone de datos para Trinidad y Tobago, se decidió no incluirlos en el análisis debido a inconsistencias y problemas en la calidad de la información.

⁶ Las bases de datos empleadas consisten en muestras aleatorias tomadas de los censos oficiales de cada uno de estos países y abarcan entre un 2% y un 10% de la población total.

⁷ La armonización engloba diversas unidades administrativas que pueden variar entre países. En ese sentido, *geolev1* puede corresponder a provincias, regiones o departamentos, mientras que *geolev2* podría representar departamentos, distritos o municipalidades.

⁸ Este fenómeno ha sido descrito en la literatura, incluyendo trabajos de Berniell et al. (2016, 2021).

incluso la revirtió en las décadas más recientes.⁹¹⁰ En tercer lugar, se observa una disminución en la proporción de personas en pareja a lo largo del tiempo —fenómeno conocido como *unmating*— y, en línea con ello, se documenta que las mujeres que forman pareja presentan, en promedio, un nivel educativo menor que aquellas que no lo hacen. Esto resulta en una reversión de la brecha de género educativa más pronunciada entre las mujeres solteras.

Tabla 1. Estadísticas descriptivas, por estado de emparejamiento

Panel A. En pareja						
Medias muestrales	1960		1990		2010	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Share (%)	64.92	64.06	60.13	61.31	52.88	51.42
Edad	34.16	35.13	34.20	34.99	35.11	35.71
Años de escolaridad	3.08	3.53	6.72	7.24	10.25	10.14
PRII (%)	78.14	76.02	37.37	32.45	13.42	12.10
PRIC (%)	18.30	18.84	40.73	44.26	34.21	37.02
SEC (%)	3.14	3.49	18.24	17.88	37.27	38.06
UNIC (%)	0.42	1.65	3.67	5.41	15.11	12.82
Obs.	1,424,584	1,421,867	3,954,365	3,853,054	3,119,801	2,920,138
Panel B. Solteros						
Medias muestrales	1960		1990		2010	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Share (%)	35.08	35.94	39.87	38.69	47.12	48.58
Edad	33.70	32.30	33.14	31.99	33.64	32.92
Años de escolaridad	3.62	3.69	7.56	7.58	10.93	10.38
PRII (%)	72.30	74.13	31.20	30.12	11.42	12.47
PRIC (%)	21.99	20.30	38.72	42.45	28.18	32.42
SEC (%)	4.85	3.77	24.32	21.41	41.37	40.77
UNIC (%)	0.87	1.80	5.75	6.02	19.03	14.33
Obs.	528,257	519,202	2,012,951	1,818,742	2,198,090	1,966,669

Nota: La tabla presenta el promedio regional de medias muestrales de edad y educación, así como también el total de observaciones para cada década del censo. Los niveles educativos presentados en la tabla son Primaria Incompleta (PRII), Primaria Completa (PRIC), Secundaria Completa (SEC) y Universitario Completo (UNIC). Se divide la muestra entre individuos solteros y aquellos en pareja. Es importante destacar que los países incluidos en cada década pueden variar de acuerdo con la disponibilidad de datos para cada década y país.

Fuente: IPUMS (2020)

La Figura 1 complementa las estadísticas presentadas en la Tabla 1 al documentar la distribución espacial del nivel educativo en la región. En particular, muestra el promedio de años de escolaridad por unidad administrativa de segundo nivel para aquellos países con información disponible en la década de 2000.¹¹ Los resultados evidencian una marcada heterogeneidad tanto entre países —con Argentina, Chile y

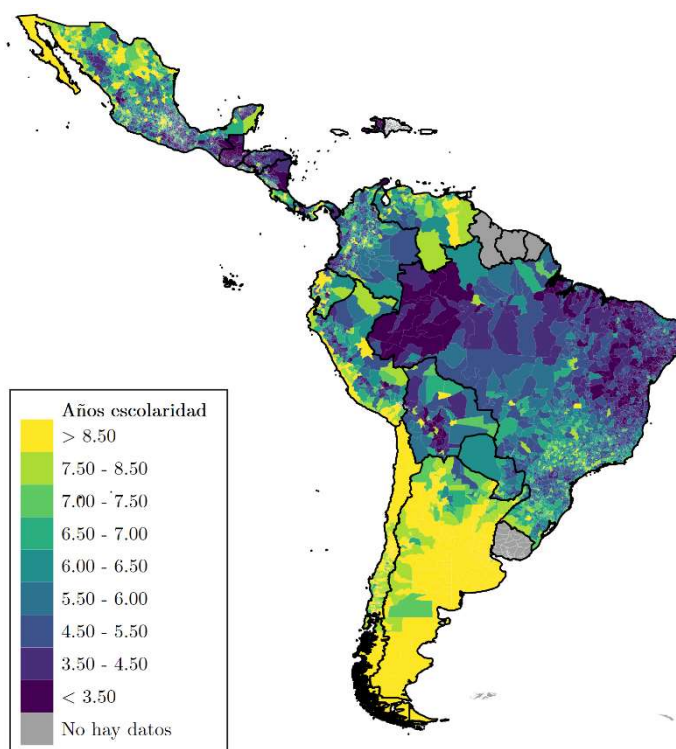
⁹ Este fenómeno ha sido descrito en la literatura, incluyendo trabajos de Berniell et al. (2016, 2021).

¹⁰ A pesar de las variaciones en los países analizados y las tendencias cambiantes, el análisis por país confirma las mismas conclusiones. La [Figura A1 del Apéndice](#) ilustra un crecimiento significativo en la educación en todos los países de América Latina y el Caribe en los últimos 100 años, tanto para hombres como, de manera más pronunciada, para mujeres.

¹¹ Se presentan los datos de esta década dada la cantidad de países con información disponible. Si bien se ha evidenciado una gran expansión educativa en los últimos años, los datos disponibles más recientes muestran que los países han mantenido la heterogeneidad intra-regional.

Costa Rica entre los de mayor escolaridad promedio— como dentro de cada uno. En países como México y El Salvador, por ejemplo, las unidades administrativas más educadas registran promedios de años de educación formal hasta diez u once veces superiores a los observados en las unidades menos educadas.

Figura 1. Años de escolaridad promedio, por unidad administrativa de segundo orden. Países de América Latina y el Caribe, censos década 2000

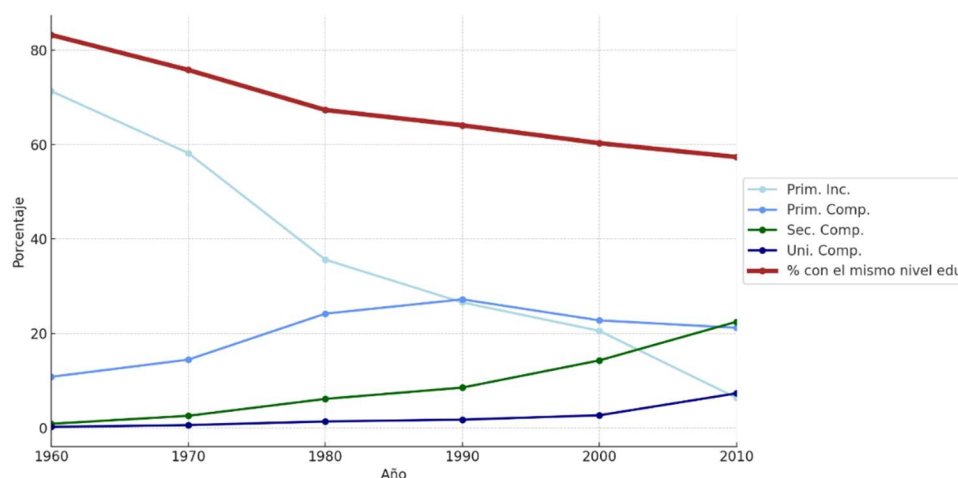


Nota: esta figura presenta los años de educación promedio para cada unidad administrativa de segundo orden para aquellos países que disponen de censos poblacionales en la década del 2000.

Fuente: IPUMS-I (2020).

La evolución del nivel educativo promedio se refleja también en la transformación de la composición educativa de las parejas a lo largo del tiempo. La Figura 2 presenta, para el promedio de países de América Latina y el Caribe, la proporción total de parejas homogámicas en términos educativos —esto es, parejas en las que ambos integrantes alcanzan el mismo nivel educativo—, así como su desagregación por nivel. La evolución correspondiente para cada país se reporta en el Apéndice (Figura A2). Los resultados indican que los cambios en las distribuciones marginales de educación en la región han estado acompañados por variaciones en la incidencia de la homogamia, tanto dentro de cada nivel como en el total de parejas. En particular, la proporción de parejas en las que ambos miembros tienen primaria incompleta disminuye significativamente entre 1960 y 2010, mientras que la homogamia en los tres niveles educativos superiores aumenta. Sin embargo, la reducción del grupo de parejas con primaria incompleta no es compensada plenamente por el incremento en los niveles restantes, de modo que, en conjunto, se observa una caída en la proporción total de parejas con el mismo nivel educativo a lo largo del período.

Figura 2. Porcentaje de parejas por nivel educativo. Promedio 21 países de América Latina y el Caribe, por década de censo



Nota: esta figura presenta el promedio simple regional del porcentaje de parejas con el mismo nivel educativo, desglosado por nivel educativo y para el total de la población. Los países incluidos pueden variar según la disponibilidad de censos.

Fuente: IPUMS-I (2020).

Las estadísticas presentadas sugieren, entonces, una relación en buena medida mecánica entre los cambios en las distribuciones educativas y los patrones observados de homogamia. En particular, la marcada reducción en la proporción de individuos con primaria incompleta parece haberse traducido en una disminución de las parejas ubicadas en ese nivel y, en consecuencia, en las tendencias agregadas de homogamia educativa. Esta evidencia motiva la siguiente pregunta: ¿cómo habrían evolucionado los patrones de emparejamiento si las distribuciones marginales de educación se hubieran mantenido constantes? En términos analíticos, el objetivo es descomponer los cambios en el emparejamiento en dos componentes: aquellos que responden a variaciones “mecánicas” derivadas de la composición educativa de la población y aquellos asociados a cambios en la tecnología —o, equivalentemente, en los “precios”— de emparejamiento. Este segundo componente es el más informativo para comprender el vínculo entre emparejamiento selectivo y desigualdad de ingresos.

Asimismo, las estadísticas descriptivas muestran una caída sostenida en la proporción de personas en pareja en la región, junto con una heterogeneidad educativa considerable entre regiones dentro de cada país. Por estos motivos, el análisis incorpora explícitamente dos dimensiones adicionales: la selección asociada a la decisión de formar pareja y la dimensión geográfica del emparejamiento, con el fin de caracterizar de manera más precisa el nivel y la evolución del emparejamiento selectivo.

3. Metodología

Esta sección presenta las métricas utilizadas para medir el emparejamiento selectivo y discute consideraciones metodológicas relevantes para su medición y comparación en el tiempo. La literatura ha propuesto y aplicado múltiples indicadores; sin embargo, algunos de ellos son sensibles a cambios en las distribuciones marginales de educación —derivados, por ejemplo, de procesos de expansión educativa y de variaciones en las brechas de género—, lo que puede dificultar la comparación intertemporal e internacional cuando el objetivo es identificar cambios asociados a la tecnología de emparejamiento. Para organizar la discusión, se agrupan las métricas en tres conjuntos (“generaciones”) según el tratamiento que otorgan a estos cambios composicionales y su capacidad para aislar el componente atribuible a variaciones en la tecnología de emparejamiento. El detalle de definiciones, ejemplos y ejercicios complementarios se reporta en el Apéndice.

3.1. Métricas tradicionales y limitaciones frente a cambios en las distribuciones educativas

Un primer conjunto de indicadores —utilizado ampliamente en la literatura inicial— cuantifica el emparejamiento selectivo educativo sin distinguir entre cambios inducidos por variaciones en las distribuciones marginales de educación y cambios atribuibles a la tecnología de emparejamiento. En esta familia se incluyen medidas descriptivas como la proporción de parejas homogámicas (o definiciones afines, por ejemplo, permitiendo “un nivel de diferencia”), así como variantes normalizadas.¹² Si bien estas métricas son intuitivas, pueden responder mecánicamente a cambios en la composición educativa y, además, dependen del esquema de categorización utilizado para agrupar la educación (Gihleb & Lang, 2020). También se han utilizado indicadores basados en regresiones lineales de la educación de un cónyuge sobre la del otro, típicamente estimadas por Mínimos Cuadrados Ordinarios, en aplicaciones para Estados Unidos y para Argentina, entre otros países (Gabrielli & Serio, 2017; González & Dip, 2022; Greenwood et al., 2014). Sin embargo, como discuten Eika et al. (2019), este coeficiente puede ser poco adecuado para estudiar la evolución del ES, dado que puede ser sensible tanto a cambios en las distribuciones educativas como a la forma de especificar la regresión (por ejemplo, el orden entre variable dependiente e independiente); el Apéndice II ilustra este punto con un ejemplo para Argentina. Finalmente, una medida emparentada con esta aproximación es la correlación lineal de Pearson, utilizada en trabajos aplicados al ES y sus implicancias distributivas (CEDLAS & World Bank, 2024; Fernandez et al., 2005; Kremer, 1997). Si bien más adelante se muestra que la correlación de Pearson puede comportarse de manera robusta en ciertos contextos, su interpretación

¹² Shorrocks (1978) propuso un índice de movilidad estrechamente relacionado con el este indicador de homogamia, el cual se denominó “Traza Normalizada”. Este se define como $NT = [n - \text{traza}(A)]/(n - 1)$, donde A es la matriz de transición con n categorías y $\text{traza}(A)$ es la suma de las proporciones de transición en la diagonal principal de A . Por construcción, una IR y la Traza normalizada son insensibles a cualquier diferencia entre las matrices de transición más allá de las de las respectivas diagonales.

está anclada en una relación lineal entre años de escolaridad y, por lo tanto, no está directamente alineada con la definición ordinal del ES como asignación por posiciones relativas (rangos).

Un segundo grupo de métricas reconoce explícitamente que los cambios en las distribuciones marginales afectan la medición del ES y, para abordarlo, incorpora contrafácticos de emparejamiento aleatorio. En términos generales, estos enfoques comparan los emparejamientos observados con los esperables bajo independencia entre la educación del hombre y de la mujer, utilizando *odds ratios* y modelos *log-lineales* —desarrollados y aplicados extensamente en la sociología—, entre otros procedimientos de normalización (Bouchet-Valat, 2014; Schwartz & Mare, 2005), así como también propuestas afines en economía (Eika et al., 2019). No obstante, aun dentro de este marco, la comparabilidad intertemporal puede verse afectada cuando cambian las brechas educativas por género y/o la composición interna de los grupos educativos a lo largo del tiempo, lo que limita la capacidad de aislar el componente atribuible a cambios en la tecnología de emparejamiento (Bratsberg et al., 2023; Shen, 2019). En esta línea, Shen (2019) propone una normalización “perfecta-aleatoria” que ubica el emparejamiento observado entre esos dos límites, con el objetivo de controlar simultáneamente cambios en brechas de género y composición; sin embargo, esta métrica puede presentar dificultades cuando las distribuciones educativas de hombres y mujeres son altamente desiguales. Los detalles de estas métricas y ejercicios ilustrativos se reportan en los Apéndices III y IV.

En síntesis, si bien las medidas de primera y segunda generación han sido ampliamente utilizadas y han contribuido a caracterizar el fenómeno, su sensibilidad a transformaciones composicionales puede dificultar inferencias sobre cambios asociados a la tecnología de emparejamiento en comparaciones intertemporales e internacionales. Esto motiva adoptar métricas basadas en posiciones relativas (rangos/cuantiles), que ofrecen mejores propiedades para estudiar la evolución del emparejamiento selectivo cuando las distribuciones educativas cambian con el tiempo.

3.2. Indicadores robustos: correlación de rangos y medidas por cuantiles

Las métricas que mejor se alinean con la definición conceptual de ES —emparejamiento “ordenado” entre posiciones relativas en las distribuciones educativas— son aquellas basadas en rangos y, en particular, las correlaciones de rangos. En un escenario de emparejamiento selectivo perfecto, los individuos más educados de cada género se emparejan entre sí, luego los segundos más educados, y así sucesivamente; este criterio es ordinal y no requiere suponer una relación lineal en años de escolaridad. En consecuencia, en el análisis se utilizan como referencia medidas de correlación de rangos (por ejemplo, Spearman y Kendall), que han mostrado un desempeño robusto en ejercicios de simulación y aplicaciones empíricas recientes, incluso en presencia de empates educativos frecuentes (Gihleb & Lang, 2020)¹³; el Apéndice V presenta evidencia complementaria en esta línea para el caso de Argentina. Asimismo, dado que la expansión educativa puede generar diferencias sistemáticas por cohorte dentro de un mismo censo y

¹³ En el trabajo de Gihleb & Lang (2020) se presenta en detalle cada uno de los indicadores utilizados. Alternativamente, el trabajo de Agresti (2010) presenta una revisión detallada de los indicadores de correlación de variables categóricas.

que los cónyuges tienden a emparejarse con edades similares, es relevante considerar *rankings* definidos por cohorte y género, de modo que la posición relativa educativa se interprete de manera consistente a lo largo del tiempo (Handy, 2021).

Sin embargo, las correlaciones agregadas no permiten, por sí solas, caracterizar cómo varía el emparejamiento a lo largo de la distribución educativa —un aspecto relevante para estudiar sus implicancias distributivas—. Para capturar esa heterogeneidad, este trabajo propone y utiliza una métrica basada en cuantiles educativos, inspirada en enfoques recientes que miden el emparejamiento relativo entre posiciones de la distribución. En particular, se retoma la lógica de indicadores cuantílicos que comparan las probabilidades observadas de emparejamiento entre cuantiles con las probabilidades esperadas bajo emparejamiento aleatorio; este enfoque, aplicado originalmente a variables socioeconómicas continuas, presenta ventajas para comparaciones intertemporales al fijar por construcción el tamaño relativo de los grupos (Bratsberg et al., 2023) y se vincula con propuestas previas que introducen normalizaciones respecto de contrafácticos aleatorios (Eika et al., 2019).

3.3. Indicador Cuantílico Educativo Local

Siguiendo la lógica de emparejamiento por cuantiles propuesta por Bratsberg et al. (2023), esta subsección introduce el Indicador Cuantílico Educativo Local (ICEL). La diferencia central respecto de su formulación original es que, en lugar de definir cuantiles sobre variables originalmente continuas asociadas a ingresos o riqueza, el ICEL utiliza cuantiles educativos contruidos a partir de una medida educativa transformada (definida más adelante) que busca aislar el componente de tendencia vinculado a la expansión educativa y facilitar comparaciones intertemporales.

Sea $CE_h(CE_m)$ el cuantil educativo del hombre (mujer). Para cada par (ce_h, ce_m) , el ICEL se define como:

$$\text{ICEL}(ce_h, ce_m) = \frac{P(CE_h = ce_h, CE_m = ce_m)}{P(CE_h = ce_h) \cdot P(CE_m = ce_m)}.$$

El indicador compara la probabilidad observada de emparejamiento entre cuantiles educativos con la probabilidad esperada bajo emparejamiento aleatorio en términos educativos, entendiendo este último en sentido estadístico como independencia entre CE_h y CE_m . En particular, si se trabaja con cuantiles de tamaño fijo, el contrafactual aleatorio queda determinado por construcción: con quintiles, $P(CE_h = ce_h) = 0.2$, por lo que la probabilidad esperada en cada celda es 0.04; con deciles, es 0.01.¹⁴

Para resumir el emparejamiento selectivo “homogámico” en términos de posiciones relativas, se calcula un promedio ponderado de los valores del ICEL en la diagonal principal, $\text{ICEL}(ce, ce)$. Un valor mayor (menor) que 1 indica emparejamiento selectivo positivo (negativo) para ese cuantil.

¹⁴ Recuérdese que la simulación de emparejamiento aleatorio lo es en un sentido estadístico, es decir, asume que E_o y E_a son variables aleatorias independientes, entonces la probabilidad aleatoria es $P(E_h = e_h) * P(E_m = e_m)$. En caso de los deciles, será $0.1 * 0.1$.

Transformación de la variable educativa y construcción de cuantiles

En microdatos censales, los años de escolaridad presentan dos rasgos que dificultan definir cuantiles directamente sobre la variable observada. Primero, la educación está acotada entre 1 y 18 años y exhibe fuertes concentraciones en valores asociados al cumplimiento de credenciales (por ejemplo, finalización de primaria o secundaria), lo que genera una alta frecuencia de empates (“*ties*”) y puede inducir asignaciones inestables al construir cuantiles. Segundo, en América Latina y el Caribe existe una tendencia secular creciente de la educación por cohorte, reflejo de la expansión educativa: individuos más jóvenes tienden a ubicarse en niveles de escolaridad más altos dentro de un mismo censo. Dado que las parejas suelen formarse entre personas de edades similares, no ajustar por esta tendencia puede contaminar la medición del emparejamiento al confundir similitud por cohorte con selección en posiciones relativas.

Para definir cuantiles comparables entre cohortes dentro de cada período, se sigue el enfoque de Halliday et al. (2021). En particular, para cada país, década censal y género, se estima una regresión de años de educación sobre efectos fijos de cohorte (año de nacimiento) y se construyen los cuantiles a partir de los residuos. Esta transformación produce una medida educativa “neta” de la tendencia por cohorte y permite interpretar los cuantiles como posiciones relativas dentro de la distribución educativa relevante para cada período. Los detalles de especificación y de implementación se presentan en el Apéndice VI.

3.4. Extensiones: selección hacia el emparejamiento y dimensión geográfica

Indicador Cuantílico Educativo Global (ICEG)

La mayoría de las métricas de emparejamiento selectivo se calculan condicionando a estar en pareja. Si la decisión de emparejarse no es aleatoria y está asociada a características observables —en particular, educación—, la evolución del ES medida en el universo de parejas puede verse afectada por cambios en quiénes integran ese universo (un aspecto especialmente relevante en la región dada la caída de la proporción de personas en pareja, *unmating*).

Para incorporar esta dimensión, se define el Indicador Cuantílico Educativo Global (ICEG) de manera análoga al ICEL, con la diferencia de que los cuantiles educativos se construyen sobre el universo de potenciales parejas (individuos en el rango etario de análisis), y no únicamente sobre quienes efectivamente cohabitan. De este modo, el ICEG resume conjuntamente (i) los patrones de asignación educativa dentro de las parejas y (ii) la selección asociada a la formación de pareja.¹⁵

Ejemplo 1. Supóngase que, dentro del grupo que efectivamente se empareja, el patrón de asignación educativa entre cónyuges permanece estable a lo largo del tiempo. Aun así, si con el tiempo aumenta la proporción de individuos —por ejemplo, mujeres con

¹⁵ El escenario contrafactual, en este caso, hace referencia a un caso teórico en el cual los individuos de cada decil podrían formar pareja con individuos del otro género perteneciente al mismo decil. En la práctica, el número de hombres y mujeres en cada muestra no coincide, con lo cual quedarían individuos sin encontrar pareja.

mayor educación— que permanecen fuera del mercado de parejas (o retrasan la cohabitación), entonces las métricas condicionadas a estar en pareja pueden sugerir cambios en el ES simplemente por la recomposición del universo emparejado. Al construir los cuantiles sobre el universo de potenciales parejas, el ICEG permite que esa dimensión quede incorporada explícitamente en la medición.

Indicador Cuantílico Educativo Geográfico (ICE Geo)

Los niveles de emparejamiento selectivo y sus tendencias han sido estudiados, en general, a nivel nacional. El emparejamiento selectivo y los niveles educativos promedio pueden variar sustantivamente entre regiones. Esta heterogeneidad es relevante en América Latina y el Caribe dada la presencia de desigualdades espaciales marcadas, baja movilidad interna y mercados matrimoniales con fuerte componente local (De La Mata et al., 2022).

Para incorporar la dimensión espacial, se propone el Indicador Cuantílico Educativo Geográfico (ICE Geo), que mide la intensidad del emparejamiento por cuantiles dentro de unidades subnacionales. La referencia contrafactual es un escenario en el que (i) el emparejamiento es aleatorio y, además, (ii) la distribución educativa local coincide con la distribución nacional. Esta extensión permite comparar regiones en una escala común y distinguir si la selección se concentra en ciertos cuantiles educativos en contextos territoriales con distinto capital educativo promedio, con potenciales implicancias diferentes para la desigualdad agregada.

Ejemplo 2. Considérense dos regiones con un grado similar de emparejamiento “dentro de la región” (por ejemplo, alta selección en el quintil superior). Si la Región A es, en promedio, mucho más educada que la Región B, entonces un “quintil alto” en A corresponde a niveles absolutos de escolaridad sistemáticamente mayores que un “quintil alto” en B. En ese caso, un indicador puramente local podría sugerir patrones comparables, aun cuando el significado distributivo sea distinto. Al normalizar con una referencia que iguala la distribución local a la nacional, el ICE Geo permite separar (a) la intensidad del emparejamiento por posición relativa de (b) el nivel educativo del contexto territorial en el que ocurre.

3.5. Validación y comparación con indicadores de referencia

Esta subsección evalúa el desempeño del Indicador Cuantílico Educativo Local en dos dimensiones. Primero, se analiza si el ICEL es robusto a cambios en las distribuciones marginales de educación. Segundo, se examina su consistencia con una medida ordinal estándar de emparejamiento selectivo, el coeficiente de correlación de rangos de Spearman.

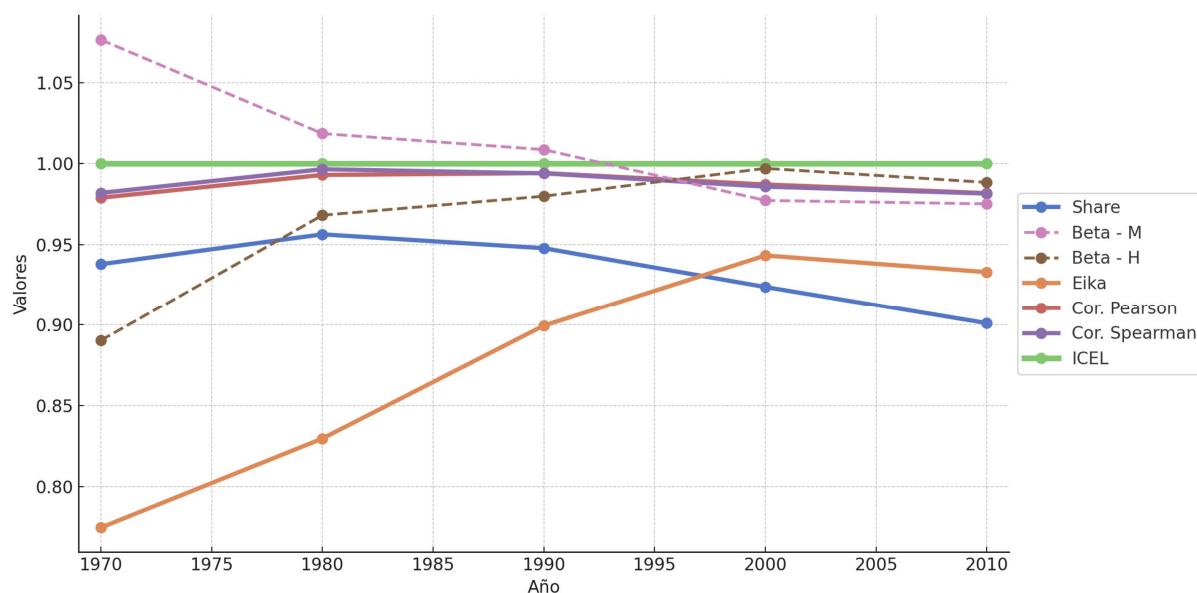
(i) Robustez ante cambios composicionales.

Para evaluar la sensibilidad de distintas métricas a variaciones en las distribuciones marginales, se implementa un ejercicio de microsimulación con datos de Argentina. En cada década, se construye un escenario de emparejamiento selectivo perfecto: la mujer con mayor educación se empareja con el hombre más educado, luego la segunda con el segundo, y así sucesivamente. Por construcción, el patrón de emparejamiento permanece constante, mientras que las distribuciones educativas por género evolucionan en línea con los datos censales. En este contexto, un indicador apropiado para comparaciones intertemporales debería permanecer estable (o, al menos, no exhibir tendencias espurias) cuando el único cambio proviene de la composición educativa.

En la comparación se incluyen medidas representativas de las generaciones discutidas previamente: (a) el *share* de homogamia; (b) coeficientes de regresiones MCO alternando el orden de la variable dependiente (Beta-M y Beta-H); (c) el indicador de Eika et al. (2019); y (d) medidas de correlación (Pearson y Spearman), junto con el ICEL. Dado que la simulación de ES perfecto genera numerosas celdas con frecuencia nula, no se implementan aquí modelos log-lineales al estilo de Schwartz y Mare (2005) (véanse detalles y alternativas en el Apéndice).

La Figura 3 resume los resultados: mientras varios indicadores tradicionales se desvían de manera apreciable del valor consistente con ES perfecto, las medidas de tercera generación (Spearman y el ICEL) se mantienen cercanas al *benchmark*, lo que sugiere mayor robustez frente a cambios en las distribuciones educativas.

Figura 3. Resultados de cada indicador bajo ES perfecto.



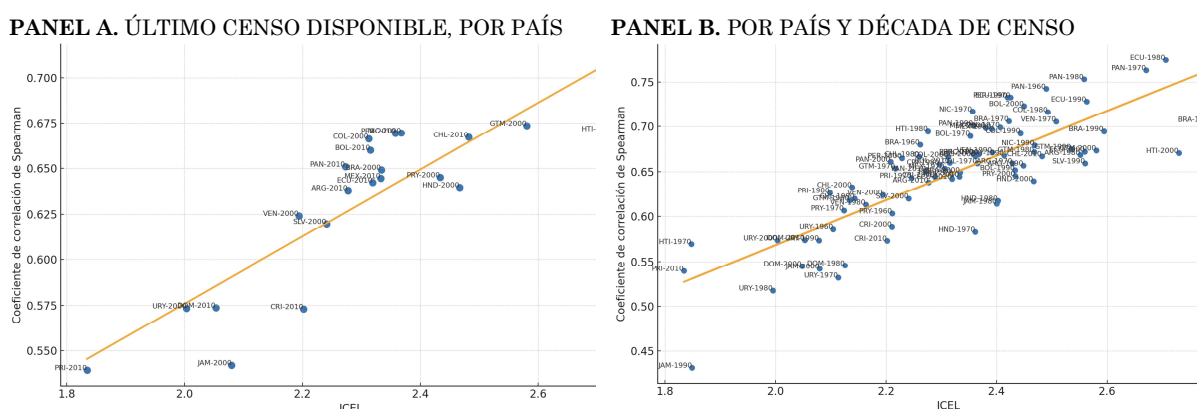
Nota: La figura muestra la evolución del emparejamiento selectivo en un escenario simulado de ES perfecto, a través de la estimación de diversos indicadores. Estos incluyen: el porcentaje de parejas con el mismo nivel educativo (*share*), los coeficientes de regresión donde la educación de la mujer es la variable independiente (Beta-M) y aquellos donde la educación del hombre es la independiente (Beta-H), el indicador propuesto por Eika et al. (2019), los coeficientes de correlación lineal de Pearson y de rangos de Spearman, y, por último, el ICEL.

Fuente: elaboración propia en base a IPUMS (2020).

(ii) Coherencia con métricas ordinales: comparación con Spearman

El ICEL se construye a partir del promedio ponderado de ES “en la diagonal” (emparejamientos dentro del mismo cuantil), por lo que podría preocupar que ignore parte de la estructura fuera de la diagonal. Para evaluar si, pese a esta característica, captura adecuadamente variación agregada del ES, se compara el ICEL con el coeficiente de Spearman. La Figura 4 muestra una asociación estrecha entre ambas métricas, tanto en el último censo disponible por país (Panel A) como en el conjunto de países y décadas (Panel B). Para el último dato disponible, la correlación entre ICEL y Spearman es 0.85, lo que respalda que el ICEL contiene información similar a una medida ordinal estándar, con la ventaja adicional de ofrecer una interpretación directa en términos de razones de probabilidades relativas al emparejamiento aleatorio.

Figura 4. Valores del ICEL y del coeficiente de Spearman.



Nota: La figura muestra la relación entre el ICEL y el coeficiente de correlación de Spearman. La línea de tendencia representa la línea de tendencia en cada gráfico. En el Panel A, se presenta el valor de cada indicador para el último censo disponible, mientras que, en el Panel B, se muestran las métricas correspondientes a todos los años analizados.

Fuente: elaboración propia en base a IPUMS (2020).

En síntesis, las Figuras 3 y 4 sugieren que el ICEL (i) presenta buen desempeño en un entorno diseñado para aislar cambios composicionales y (ii) es consistente con una medida ordinal consolidada como Spearman, aun cuando su construcción enfatiza los emparejamientos dentro de cuantiles equivalentes. Esta evidencia motiva su uso en el análisis empírico que sigue y sus extensiones (ICEG e ICE Geo), que incorporan selección en la formación de pareja y heterogeneidad espacial.

4. Resultados

Una vez definidos el Indicador Cuantílico Educativo Local y el coeficiente de correlación de rangos de Spearman como métricas con buenas propiedades para medir el nivel y la evolución del emparejamiento selectivo bajo cambios en las distribuciones educativas, esta sección documenta (i) los niveles de ES en América Latina y el Caribe, (ii) su evolución temporal, (iii) la heterogeneidad por quintiles, y (iv) los resultados de las extensiones metodológicas propuestas: el indicador global y el indicador geográfico.

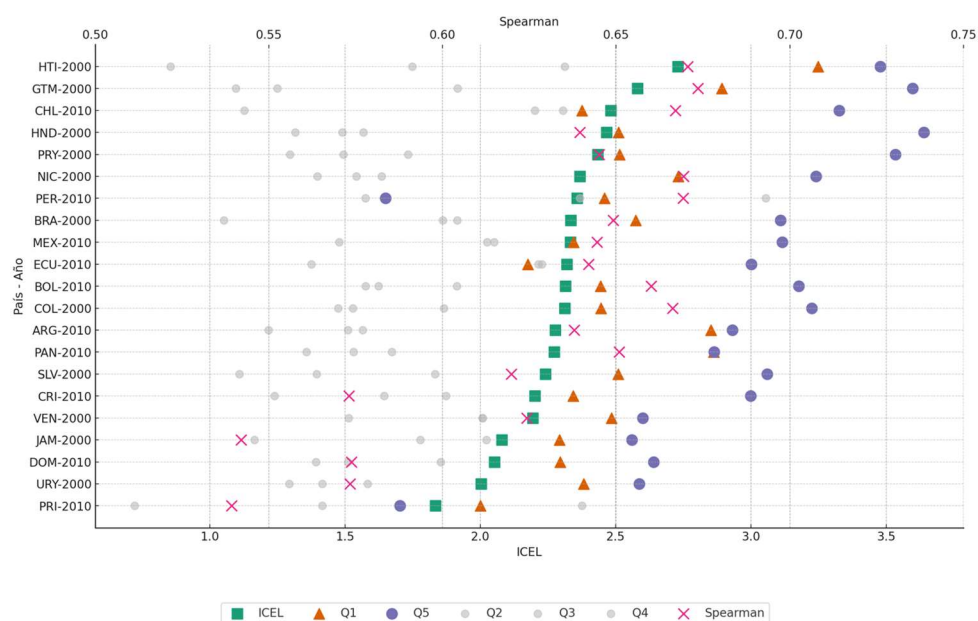
4.1. Emparejamiento selectivo en la región

La Figura 5 presenta, para el último censo disponible de cada país, el ICEL agregado, el ICEL por quintil y el coeficiente de Spearman. Los resultados muestran evidencia de emparejamiento selectivo positivo en todos los países: en promedio, la probabilidad de emparejarse con una persona del mismo quintil educativo es sustancialmente mayor que la que surgiría bajo emparejamiento aleatorio.

La Figura también revela una heterogeneidad marcada tanto entre países como a lo largo de la distribución educativa. En el promedio simple regional, los individuos son aproximadamente 2.3 veces más propensos a emparejarse dentro de su mismo quintil que bajo aleatoriedad. Sin embargo, este valor se ubica cerca de 2 en algunos países y alrededor de 3 en otros (por ejemplo, países del Caribe tienden a ubicarse en el extremo inferior, mientras que varios países de Centroamérica se ubican en el superior). Asimismo, la intensidad del ES difiere entre extremos: el ICEL medio en Q1 ronda 2.5, mientras que en Q5 se aproxima a 3, lo que sugiere que la selección es, en promedio, más intensa en la parte alta de la distribución educativa.

Por último, el ordenamiento de países según Spearman (promedio cercano a 0.63) es, en términos generales, consistente con el patrón que emerge del ICEL: los países con mayor selección según el ICEL tienden también a mostrar correlaciones de rangos más elevadas.

Figura 5. Resultados del ICEL: último censo disponible, por país.



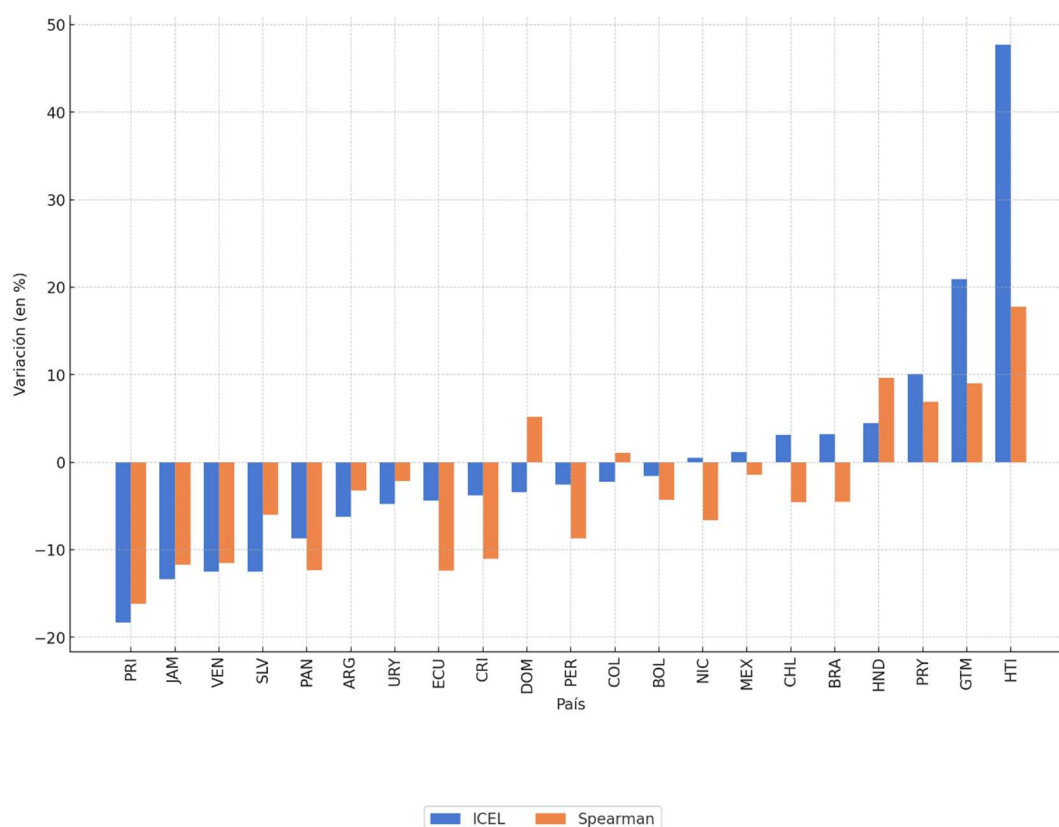
Nota: La figura muestra el valor agregado del coeficiente de Spearman y del ICEL agregado junto con los valores del ICEL para cada uno de los cuantiles de la distribución educativa. Los quintiles 2 a 4 se representan en color gris. Los valores del ICEL se ubican en el eje inferior, mientras que en el eje superior se presentan los correspondientes al coeficiente de correlación de rangos de Spearman. Cada punto refleja los datos disponibles para los distintos censos en cada país.

Fuente: Elaboración propia con base a IPUMS (2020).

4.2. Evolución del ES en la región

La Figura 6 resume la variación porcentual del ICEL y del coeficiente de Spearman entre el primer y el último registro disponible por país (las trayectorias completas por país y década se presentan en la Figura A3 del Apéndice). En el agregado, los resultados apuntan a una reducción del ES en una parte importante de la región: según el ICEL, 13 de 21 países exhiben caídas entre periodos; utilizando Spearman, las caídas se observan en 15 de 21 países y, en promedio, son más pronunciadas.

Figura 6. Evolución del ES en América Latina y el Caribe



Nota: El gráfico presenta la variación porcentual en el ICEL agregado y en el coeficiente de Spearman entre el primer y el último registro disponible en cada país. Los rangos temporales considerados son los siguientes: Argentina (1970-2010), Bolivia (1970-2010), Brasil (1960-2000), Chile (1970-2010), Colombia (1970-2000), Costa Rica (1970-2010), República Dominicana (1970-2010), Ecuador (1970-2010), El Salvador (1970-2000), Guatemala (1960-2000), Haití (1970-2000), Honduras (1970-2000), Jamaica (1970-2000), México (1970-2010), Nicaragua (1970-2000), Panamá (1960-2010), Paraguay (1960-2000), Perú (1970-2010), Puerto Rico (1970-2010), Uruguay (1960-2000) y Venezuela (1970-2000).

Fuente: elaboración propia en base a IPUMS (2020)

La Figura 6 también sugiere que existe heterogeneidad relevante en la evolución: algunos países presentan descensos cercanos al 10% según ambas métricas, mientras que otros muestran incrementos apreciables. En un subconjunto de casos, la magnitud del cambio difiere entre ICEL y Spearman, aunque las discrepancias suelen ser moderadas y no alteran el diagnóstico general sobre la dirección del cambio.

Como verificación externa, al comparar estas tendencias con estadísticas armonizadas de SEDLAC (CEDLAS & World Bank, 2024), se observan diferencias en niveles para algunos países, pero las tendencias generales se mantienen (véase Figura A4 del Apéndice).

La disminución del ES en 13 de 21 países es compatible con transformaciones demográficas y socioeconómicas documentadas para la región (Esteve et al., 2012; García & Rojas, 2004; Quilodrán, 2001). En particular, (i) la expansión educativa pudo haber ampliado y diversificado el conjunto de emparejamientos factibles; (ii) el aumento de la participación laboral femenina pudo haber modificado los ámbitos de interacción y encuentro; y (iii) la urbanización pudo haber incrementado el tamaño de los mercados matrimoniales locales y potencialmente reducido fricciones asociadas a la distancia. En los países donde el ES aumenta, el patrón observado sugiere que cambios más graduales

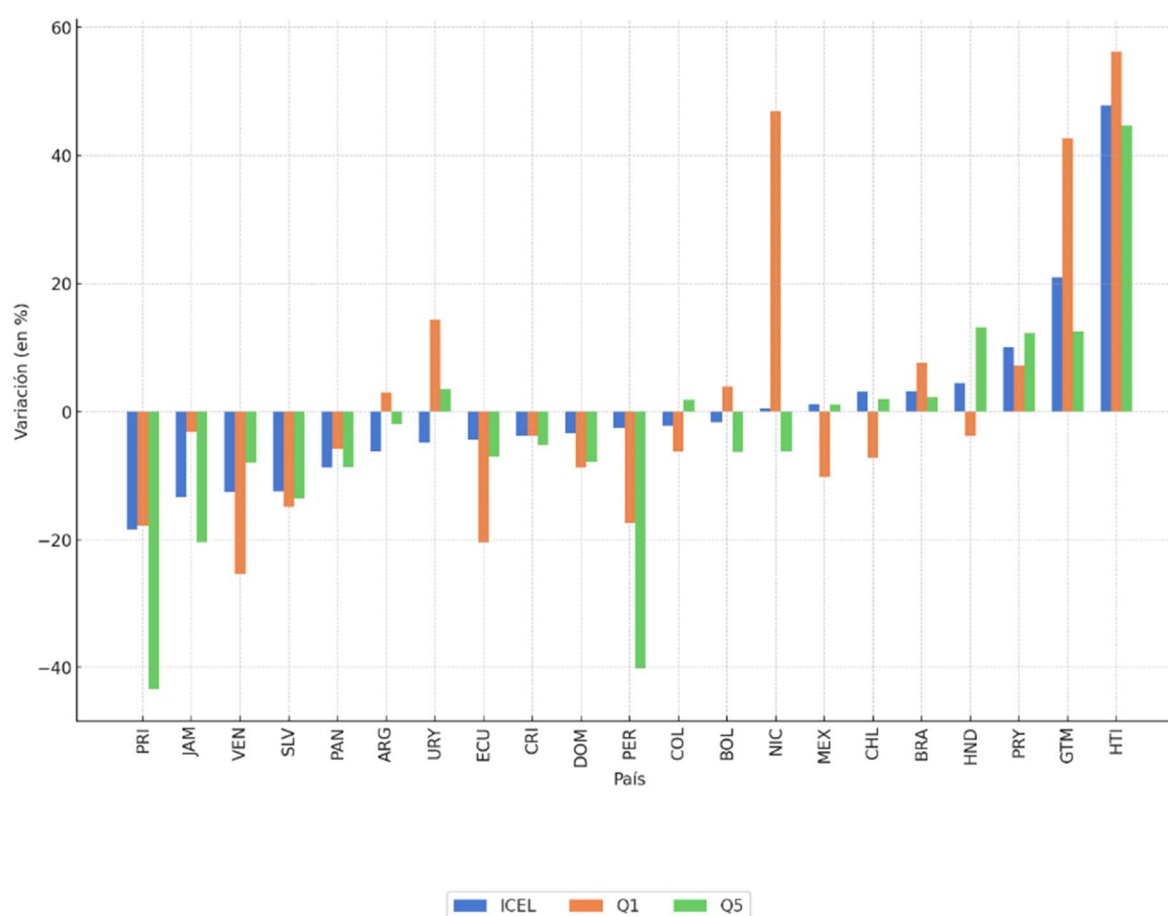
o heterogéneos en la urbanización y la expansión educativa —en especial cuando se concentran territorialmente— podrían estar asociados a una persistencia mayor de segmentación en los mercados matrimoniales. Esta interpretación se presenta como hipótesis descriptiva, más que como evidencia causal.

4.3. Evolución del ES por quintiles.

El análisis agregado puede ocultar cambios heterogéneos por nivel educativo. La Figura 7 compara la variación porcentual del ICEL agregado con la de los extremos de la distribución (Q1 y Q5) entre el primer y el último registro disponible por país (las series completas se reportan en la Figura A5 del Apéndice). En términos generales, los cambios del indicador agregado reflejan los cambios en los extremos, aunque existen excepciones relevantes en las que el agregado se mantiene relativamente estable mientras los extremos se mueven en direcciones opuestas. En particular, se observan casos donde el ES agregado cae, pero el ES en Q1 aumenta (sugiriendo una recomposición del patrón de emparejamiento hacia una mayor selección en la parte baja), y casos donde un aumento del agregado coexiste con una caída en Q1. Asimismo, hay países en los que Q1 y Q5 evolucionan en sentidos contrarios, lo cual reduce el cambio neto del indicador agregado.

De manera consistente con este patrón, los extremos parecen contribuir de forma importante a la dinámica general: la correlación entre ICEL(Q5) y el coeficiente de Spearman es elevada (≈ 0.8), mientras que para ICEL(Q1) es menor pero aún sustantiva (≈ 0.6), lo que sugiere que el componente ordinal capturado por Spearman está fuertemente alineado con la selección en los extremos de la distribución.

Figura 7. Evolución del ES por quintil educativo



Nota: El gráfico presenta la variación porcentual en el ICEL agregado y en el ICEL Q1 y Q5 entre el primer y el último registro disponible en cada país. Los rangos temporales considerados son los siguientes: Argentina (1970-2010), Bolivia (1970-2010), Brasil (1960-2000), Chile (1970-2010), Colombia (1970-2000), Costa Rica (1970-2010), República Dominicana (1970-2010), Ecuador (1970-2010), El Salvador (1970-2000), Guatemala (1960-2000), Haití (1970-2000), Honduras (1970-2000), Jamaica (1970-2000), México (1970-2010), Nicaragua (1970-2000), Panamá (1960-2010), Paraguay (1960-2000), Perú (1970-2010), Puerto Rico (1970-2010), Uruguay (1960-2000) y Venezuela (1970-2000).

Fuente: elaboración propia en base a IPUMS (2020)

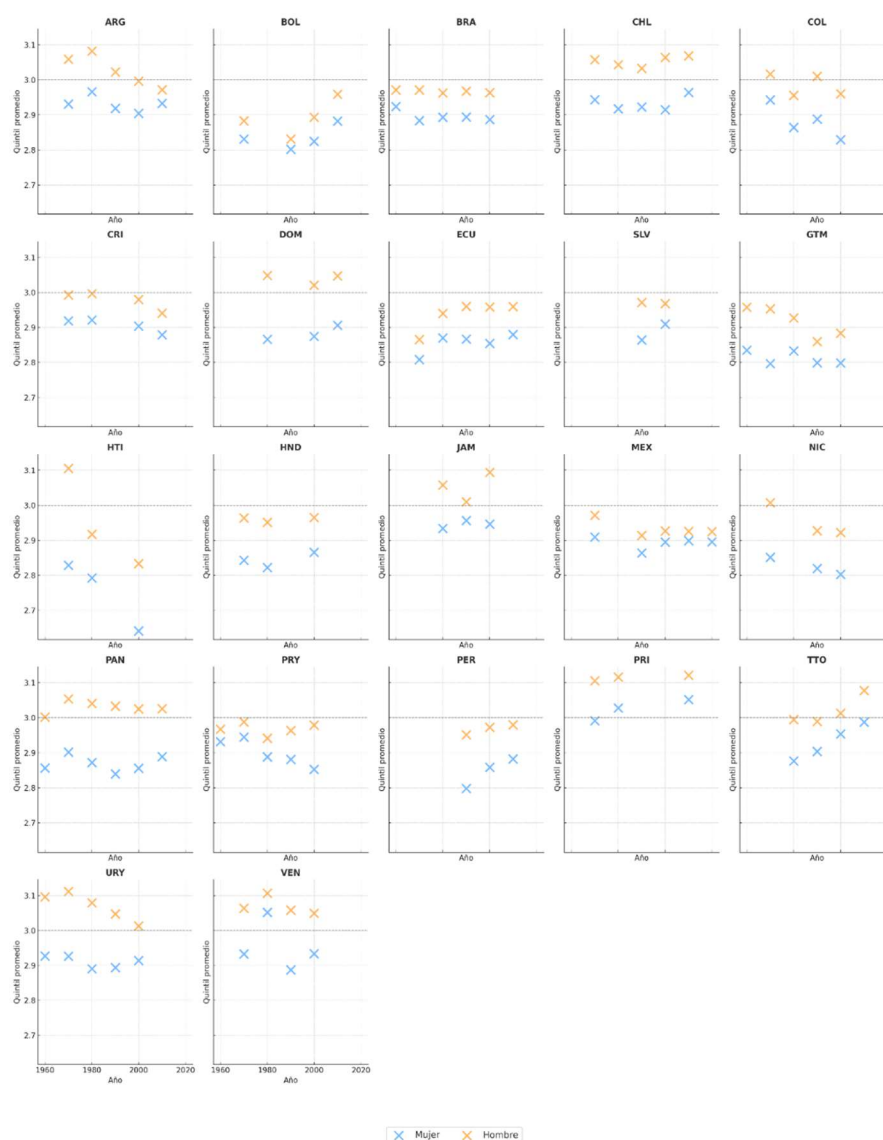
4.4. Indicador Cuantílico Educativo Global

Las métricas estándar de ES suelen calcularse condicionando a estar en pareja. Sin embargo, si la decisión de emparejarse está asociada a características observables — en particular, la educación—, la evolución medida en el universo de parejas puede verse afectada por cambios en quiénes integran ese universo, especialmente en contextos donde disminuye la proporción de población emparejada (*unmating*)¹⁶. Para incorporar esta dimensión, se propone el Indicador Cuantílico Educativo Global (ICEG), que construye los cuantiles educativos sobre el universo de potenciales parejas (individuos en el rango etario analizado) y no solo sobre quienes efectivamente cohabitan.

¹⁶ En la [Figura A6](#) se presenta evidencia que respalda la caída del porcentaje de individuos en pareja para todos los países de la región, excepto los casos de Uruguay y Paraguay.

La Figura 8 muestra evidencia consistente con selección en la formación de pareja. Si no existiera selección, el quintil promedio de quienes están en pareja debería ubicarse en torno a 3 (por construcción). En cambio, valores por debajo de 3 indican que, en promedio, quienes forman pareja se ubican en posiciones relativamente más bajas de la distribución educativa (selección negativa), mientras que valores por encima de 3 reflejan selección positiva. En la mayoría de los países, las mujeres en pareja se ubican por debajo de 3, mientras que los hombres aparecen sistemáticamente por encima de ellas (aunque no necesariamente por encima de 3). Este patrón es relevante en el contexto de la región, donde la caída en la proporción de personas en pareja puede modificar la composición del universo emparejado y, por lo tanto, afectar la comparabilidad intertemporal de métricas calculadas solo entre parejas.

Figura 8. Quintil promedio de los individuos en pareja, por país y década

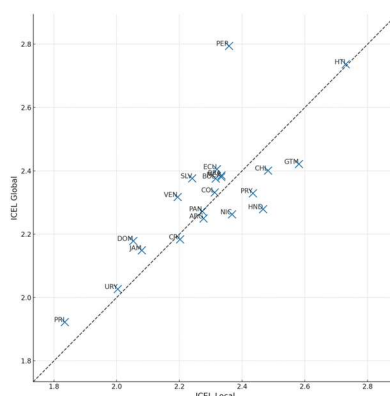


Nota: cada cruz representa el quintil promedio de la distribución educativa de los individuos que forman pareja. La línea horizontal sobre el 3 representa el valor promedio de los quintiles en caso de que no exista selección.

Fuente: elaboración propia en base a IPUMS (2020).

La Figura 9 compara el ICEG y el ICEL para el último censo disponible (las trayectorias por país se muestran en la Figura A7 del Apéndice). En conjunto, las diferencias entre ambos indicadores son, en promedio, moderadas; sin embargo, en varios países el ICEG tiende a ubicar el nivel de ES en valores levemente superiores. Este patrón es consistente con la idea de que los cambios en quiénes forman pareja —además de cómo se emparejan— pueden interactuar con el emparejamiento observado dentro del universo de parejas, y motivan considerar métricas que incorporen explícitamente la selección hacia la formación de pareja.

Figura 9. Comparación entre los niveles del ICEG e ICEL



Nota: el gráfico presenta los valores del ICEL e ICEG para el último censo disponible. La línea punteada representa una línea de 45 grados.

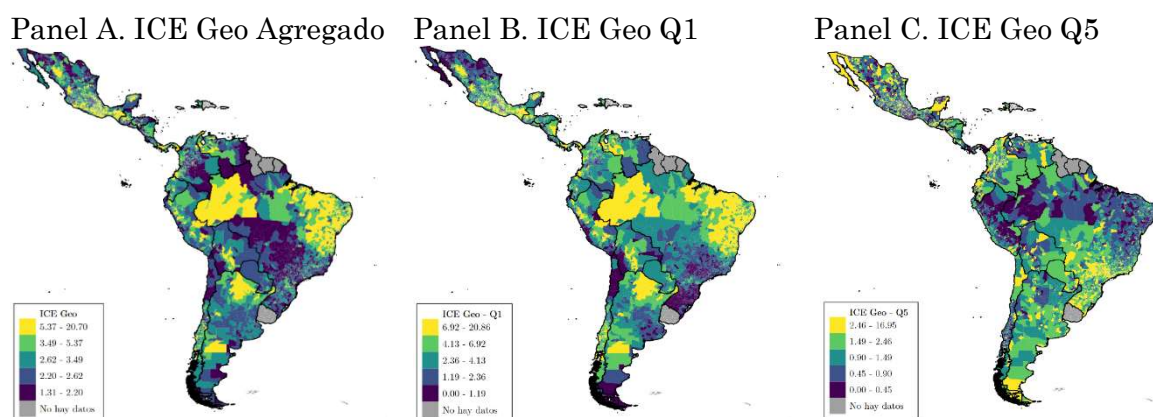
Fuente: elaboración propia en base a IPUMS (2020)

4.5. Indicador Cuantílico Educativo Geográfico

El ES y sus implicancias distributivas pueden variar significativamente en el espacio, particularmente en América Latina y el Caribe, donde las desigualdades territoriales y la segmentación de mercados matrimoniales pueden ser relevantes. La Figura 10 reporta el ICE Geo para unidades subnacionales (década de 2000)¹⁷, con un indicador agregado (Panel A) y sus versiones para Q1 y Q5 (Paneles B y C).

¹⁷ Se elige la década de 2000 por ser la de mayor cobertura de países. Aunque los mapas se construyen a nivel subnacional de segundo orden, los patrones espaciales son similares al usar agregación de primer nivel y se mantienen relativamente estables en el tiempo; no obstante, el nivel y la dispersión de los indicadores pueden variar al cambiar la unidad administrativa.

Figura 10. Indicador Cuantílico Educativo Geográfico – América Latina y el Caribe, 2000



Nota: el gráfico presenta los valores que toma el ICE Geo por unidad subnacional de segundo orden en la década de 2000. En el Panel A, se presentan los valores del ICE Geo agregado, mientras que en los paneles B y C, se presentan los valores del indicador para el Quintil 1 y 5, respectivamente.

Fuente: elaboración propia en base a IPUMS (2020)

El Panel A muestra una heterogeneidad espacial marcada en la intensidad del emparejamiento selectivo. En comparación con correlaciones subnacionales tradicionales (presentadas en la Figura A8 en el Apéndice), algunos focos territoriales de alta selección se mantienen, pero en otros contextos el patrón cambia; esto sugiere que incorporar el contrafactual y la escala de referencia nacional modifica sustantivamente la lectura del fenómeno. Los Paneles B y C evidencian, además, un patrón distributivo: en muchas áreas, la selección tiende a concentrarse diferencialmente en los extremos, y las regiones con alta selección en Q5 tienden a exhibir baja selección en Q1 (correlación aproximada -0.53).

Un aspecto clave de esta extensión es que permite comparar regiones en una escala común que incorpora el nivel educativo relativo del territorio. Por ejemplo, dos regiones pueden presentar un grado similar de emparejamiento “ordenado” entre individuos del mismo cuantil, pero si una región tiene un capital educativo promedio alto y la otra, bajo, el significado distributivo del emparejamiento —y su potencial vínculo con segregación y desigualdad— puede ser distinto. En conjunto, los resultados indican que la dimensión geográfica, al combinar composición educativa local y patrones de asignación, es un componente relevante para comprender tanto el ES como sus posibles consecuencias agregadas.

5. Determinantes del emparejamiento

El objetivo de esta sección es medir cuánto de la asociación educativa observada en las parejas puede explicarse mecánicamente por restricciones no educativas —en particular, estado de emparejamiento, edad y geografía— aun bajo un proceso de asignación aleatorio en educación. Para ello se implementa un ejercicio de simulación con microdatos censales de México en 2000 y 2020, considerando individuos de 26–60 años. En particular, para este ejercicio de simulación se amplía el rango etario (26–60) para asegurar tamaño muestral suficiente y capturar mercados matrimoniales más amplios.

Con dicho propósito, se llevará a cabo un ejercicio de simulación en el cual se partirá de suponer que el emparejamiento es completamente aleatorio en términos educativos y

se analizará cómo varían los indicadores a medida que se introducen restricciones a esta aleatorización. Se construyen cinco escenarios (R1–R5) que parten de un emparejamiento aleatorio en términos educativos e incorporan restricciones progresivas:

1. **R1. Aleatorización total:** Emparejamiento completamente aleatorio en educación entre todos los individuos (sin condicionar por estado de emparejamiento), restringiendo el número total de parejas al observado en los datos.
2. **R2. Aleatorización condicional en cohabitación:** Emparejamiento aleatorio en educación únicamente entre individuos que efectivamente cohabitan en los datos observados, manteniendo fijo el estado de emparejamiento de cada persona.
3. **R3. Aleatorización con restricción etaria:** Además de R2, se mantiene invariable la distribución de rangos etarios de las parejas. La aleatorización ocurre por separado dentro de grupos etarios (26-29, 30-39, 40-49, 50-60 años), de modo que cada pareja observada se reemplaza por dos individuos del mismo rango etario según género.
4. **R4. Aleatorización con restricción geográfica (primer orden):** Además de R3, se incorpora restricción espacial: las parejas solo pueden formarse entre individuos que residen en la misma unidad administrativa de primer orden.
5. **R5. Aleatorización con restricción geográfica (segundo orden):** Similar a R4, pero con restricción espacial más estricta: las parejas solo pueden formarse entre individuos de la misma unidad administrativa de segundo orden.

El ES se evalúa utilizando el coeficiente de Spearman y con el ICEL y el ICEG (en su versión agregada y para los quintiles extremos). La Tabla 3 reporta el promedio de los indicadores para diez aleatorizaciones por escenario.

Tabla 3: Resultados de emparejamiento selectivo para los diversos escenarios simulados

Panel A. México, 2000							
	Spearman	ICEG	ICEG Q1	ICEG Q5	ICEL	ICEL Q1	ICEL Q5
R0. Observado	0.721	2.318	2.611	3.014	2.375	2.528	3.243
R1. Aleatorización total	0.000	1.000	0.998	1.000	1.000	0.998	1.000
R2. R1 + Pareja	0.000	1.004	1.042	0.892	1.000	0.999	1.001
R3. R1 + Pareja + Grupo etario	0.088	1.028	1.054	0.907	1.040	1.017	1.012
R4. R1 + Pareja + Grupo etario + Geo 1	0.196	1.160	1.322	1.093	1.173	1.276	1.211
R5. R1 + Pareja + Grupo etario + Geo 2	0.339	1.390	1.720	1.405	1.406	1.662	1.547
Panel B. México, 2020							
	Spearman	ICEG	ICEG Q1	ICEG Q5	ICEL	ICEL Q1	ICEL Q5
R0. Observado	0.651	2.239	2.535	2.724	2.353	2.553	3.078
R1. Aleatorización total	0.000	1.001	1.000	1.002	1.001	1.001	1.002
R2. R1 + Pareja	0.000	1.017	1.000	0.806	1.001	1.002	1.000
R3. R1 + Pareja + Grupo etario	0.043	1.066	1.019	0.816	1.096	1.031	1.010
R4. R1 + Pareja + Grupo etario + Geo 1	0.124	1.172	1.275	0.935	1.208	1.291	1.149
R5. R1 + Pareja + Grupo etario + Geo 2	0.265	1.379	1.681	1.246	1.433	1.697	1.501

Nota: la tabla presenta las estimaciones de emparejamiento selectivo utilizando diferentes métricas (indicadas en cada una de las columnas). Cada fila hace referencia al escenario simulado y a las restricciones impuestas en cada una de las simulaciones. Las estimaciones de cada una de las aleatorizaciones surgen del promedio simple de las estimaciones para diez aleatorizaciones por escenario.

Fuente: IPUMS-I (2020)

Los valores presentados en la Tabla 3 revelan varios aspectos de interés. En primer lugar, el emparejamiento selectivo agregado disminuye entre 2000 y 2020 en México, de acuerdo con las medidas agregadas consideradas.

Segundo, la comparación entre R1 y R2 ilustra una diferencia conceptual entre indicadores: mientras que Spearman y el ICEL —calculados condicionando a estar en pareja— se mantienen prácticamente invariantes, el ICEG cambia al incorporar el universo de “potenciales parejas”, reflejando que la composición de quiénes se emparejan puede afectar la medición aun cuando la asignación dentro del universo emparejado sea aleatoria.

Tercero, al introducir la restricción etaria (R3) aumenta el emparejamiento medido por Spearman, consistente con que edad y educación están correlacionadas y con que los cónyuges tienden a concentrarse en rangos etarios similares. En contraste, los cambios en ICEL/ICEG son menores, en línea con que estos indicadores se construyen a partir de cuantiles definidos en base a una variable educativa a la cual se descuentan los cambios educativos vinculados con la edad de los individuos (y, por definición, son menos sensibles a desplazamientos sistemáticos de la educación por cohorte/edad).

Finalmente, las restricciones geográficas (R4–R5) incrementan sustantivamente el emparejamiento “mecánico” bajo aleatorización educativa. En particular, al imponer que las parejas se formen dentro de la misma unidad subnacional —especialmente a segundo orden—, la asociación educativa simulada pasa a explicar una fracción relevante de la correlación observada. En conjunto, este ejercicio sugiere que la dimensión espacial es un determinante cuantitativamente importante de los patrones de emparejamiento, y que una parte no despreciable del ES agregado puede emerger de la estructura local de los mercados matrimoniales, aun sin selección directa por educación.

En síntesis, la evidencia previa sugiere una heterogeneidad espacial marcada en capital educativo y en patrones de emparejamiento en la región. Consistente con ello, las simulaciones muestran que, aun bajo asignación aleatoria en educación, imponer restricciones locales —en particular geográficas— reproduce una fracción no despreciable del emparejamiento agregado observado (Tabla 3). Esto indica que la geografía emerge como un determinante relevante del ES, más allá de cambios puramente composicionales.

6. Emparejamiento selectivo y desigualdad

La literatura sobre emparejamiento selectivo ha estudiado de manera recurrente cómo cambiaría la desigualdad de ingresos de los hogares si las parejas se formaran aleatoriamente en lugar de seguir los patrones observados (Eika et al., 2019; Greenwood et al., 2014; Hryshko et al., 2017, para Estados Unidos; Funes Leal, 2015; Pereira et al., 2020, para países latinoamericanos). En estos ejercicios, la aleatorización genera pseudohogares, es decir, hogares contrafactuales contruidos al reasignar parejas manteniendo fija la población y (según el caso) ciertas características observables.

Para implementar esta idea, se han utilizado dos estrategias: el enfoque aditivo y el enfoque de imputación (Harmenberg, 2014). En el enfoque aditivo, el ingreso del pseudohogar se construye como la suma de los ingresos individuales de sus miembros reemparejados; en el enfoque de imputación, el ingreso del pseudohogar se asigna

(imputa) a partir de la distribución de ingresos observada en hogares “similares” según un conjunto de covariables, incorporando así —al menos parcialmente— que ingresos y oferta laboral pueden ser endógenos a la formación del hogar. Dado que ambos enfoques descansan sobre supuestos distintos y no son directamente comparables, aquí se adopta el enfoque aditivo por su simplicidad y transparencia (la formulación formal se presenta en el Apéndice).

En particular, se simulan los mismos escenarios de la Sección 5 (R1–R5) para México en 2000 y 2020, utilizando el ingreso laboral total de la pareja en el universo de cohabitantes. La desigualdad se mide con tres indicadores: coeficiente de Gini, la relación de ingresos entre el decil 10 y el decil 1 y el índice de Atkinson (1). La Tabla 4 resume los resultados.

Tabla 4. Indicadores de desigualdad para diversos escenarios de emparejamiento aleatorio

Panel A. México, 2000						
	Indicador			Cambios proporcionales		
	Gini	D10/D1	A(1)	Gini	D10/D1	A(1)
R0. Observado	56.4	38.0	43.9			
R1. Aleatorización total	53.1	32.2	40.1	-5.76%	-15.1%	-8.6%
R2. R1 + Pareja	53.7	32.6	40.6	-4.82%	-14.1%	-7.5%
R3. R1 + Pareja + Grupo etario	53.8	32.9	40.7	-4.68%	-13.4%	-7.2%
R4. R1 + Pareja + Grupo etario + Geo 1	54.3	34.6	41.5	-3.64%	-8.9%	-5.3%
R5. R1 + Pareja + Grupo etario + Geo 2	54.8	35.9	42.2	-2.82%	-5.5%	-3.8%
Panel B. México, 2020						
	Indicador			Cambios proporcionales		
	Gini	D10/D1	A(1)	Gini	D10/D1	A(1)
R0. Observado	46.5	25.0	33.6			
R1. Aleatorización total	43.5	20.7	30.3	-6.45%	-17.1%	-9.7%
R2. R1 + Pareja	43.4	20.5	30.1	-6.65%	-18.0%	-10.3%
R3. R1 + Pareja + Grupo etario	43.5	20.6	30.2	-6.52%	-17.5%	-10.1%
R4. R1 + Pareja + Grupo etario + Geo 1	44.3	22.1	31.1	-4.80%	-11.7%	-7.3%
R5. R1 + Pareja + Grupo etario + Geo 2	44.9	23.2	31.9	-3.41%	-7.1%	-5.0%

Nota: las primeras tres columnas de la tabla presentan métricas de desigualdad, medidas por el Coeficiente de Gini, el ratio de ingresos entre decil 10 y decil 1 y el Atkinson (1) para diferentes escenarios simulados. Cada fila hace referencia al escenario simulado y a las restricciones impuestas en cada una de las simulaciones.

Fuente: Elaboración propia en base a IPUMS (2020)

Los ejercicios muestran que el emparejamiento observado está asociado a niveles de desigualdad mayores que los que surgirían bajo aleatorización completa, aunque la magnitud del efecto depende de las restricciones incorporadas. En 2000, el Gini cae de 56.4 (observado) a 53.1 bajo aleatorización total (R1), una reducción de 5.76%; en 2020, disminuye de 46.5 a 43.5 (-6.45%). Los otros indicadores exhiben patrones consistentes: el ratio D10/D1 cae aproximadamente entre 15% y 17% y Atkinson (1) cerca de 9% bajo R1. Sin embargo, al introducir restricciones alineadas con el proceso de formación de

pareja —condicionar a estar emparejado, mantener estructura etaria y limitar el emparejamiento a mercados geográficos locales— el efecto de la aleatorización sobre la desigualdad se atenúa. Por ejemplo, en el escenario más restrictivo (R5), la reducción del Gini se reduce a -2.82% en 2000 y -3.41% en 2020, con resultados análogos para D10/D1 y Atkinson.

En suma, las simulaciones sugieren que el emparejamiento selectivo contribuye a moldear la desigualdad observada, pero que una parte relevante de esa contribución se absorbe cuando se incorporan restricciones demográficas y geográficas que estructuran los mercados matrimoniales y, por lo tanto, limitan el alcance de la aleatorización.

7. Comentarios finales

En este trabajo, documentamos el grado de emparejamiento selectivo, sus tendencias y las diferencias entre países en América Latina y el Caribe. Para ello, primero caracterizamos las métricas propuestas en la literatura, discutiendo sus limitaciones en el contexto de una región que ha experimentado una gran expansión educativa a lo largo de las décadas analizadas. La revisión de estas deja en evidencia sus limitaciones y las posibles variaciones en los resultados según la métrica empleada. Basados en esta evaluación, identificamos los indicadores más robustos y proponemos el ICEL, una versión modificada de un indicador menos sensible a los efectos de composición derivados de cambios en la distribución educativa, que permite estudiar los patrones de emparejamiento entre diferentes grupos educativos. Los resultados reflejan niveles de emparejamiento selectivo que varían entre países y presentan tendencias mayormente estables, aunque con leves descensos a partir de la década de 1990. En cuanto a los grupos educativos, los individuos de los extremos de la distribución educativa muestran mayores niveles de selectividad al formar pareja.

Reconocemos que un análisis basado únicamente en el universo de parejas formadas podría estar ignorando ciertas variaciones en los patrones de emparejamiento. Por ello, proponemos un indicador que incluye a todos los individuos que podrían formar pareja en cada país. Nuestros resultados muestran cierta selección entre quienes deciden formar una pareja, principalmente en el caso de las mujeres, y que el indicador que considera dicha selección, en general, tiende a presentar valores más altos de ES en relación al que no lo hace.

Asimismo, analizamos la dimensión espacial del emparejamiento en unidades subnacionales de segundo orden. Las heterogeneidades encontradas motivan el estudio acerca del peso relativo de dicho factor, así como de otras dimensiones, en los patrones de emparejamiento nacionales y, en consecuencia, sobre los niveles de desigualdad. Observamos que las restricciones etarias, espaciales, de estado de emparejamiento pueden explicar una porción no menor de dichos patrones.

Este análisis deja en evidencia las complejidades que reviste la medición del emparejamiento selectivo y subraya la importancia de considerar cada una de ellas para estudiar la evolución del fenómeno. A su vez, los resultados pueden ser útiles para una

agenda de investigación relevante para América Latina y el Caribe, dado que contribuyen con la producción de métricas más precisas y de amplia cobertura espacial y temporal.

Consideramos que, con un diagnóstico claro del fenómeno en la región, se puede continuar el estudio de sus determinantes y de sus implicancias sobre la desigualdad y la movilidad intergeneracional. Asimismo, entender mejor estas causas e implicancias puede contribuir al análisis de las posibles interacciones no triviales entre las políticas redistributivas y el emparejamiento selectivo, las cuales pueden potenciar o atenuar el efecto igualador de esas políticas a mediano o largo plazo.

Referencias

- Agresti, A. (2010). *Analysis of ordinal categorical data* (Vol. 656). John Wiley & Sons.
- Becker, G. S. (1974). A theory of marriage. En *Economics of the family: Marriage, children, and human capital* (pp. 299–351). University of Chicago Press.
- Berniell, L., Bonavida, C., de la Mata, D., & Schargrotsky, E. (2021). *La movilidad educativa intergeneracional en el siglo XX en América Latina y el Caribe* [workingPaper]. CAF.
<https://cafsciotea.azurewebsites.net/handle/123456789/1864>
- Berniell, L., de la Mata, D., Bernal, R., Camacho, A., Barrera-Osorio, F., Álvarez, F., Brassiolo, P., & Vargas, J. (2016). *RED 2016. Más habilidades para el trabajo y la vida: Los aportes de la familia, la escuela, el entorno y el mundo laboral*. CAF.
<https://cafsciotea.azurewebsites.net/handle/123456789/936>
- Bouchet-Valat, M. (2014). Les évolutions de l’homogamie de diplôme, de classe et d’origine sociales en France (1969-2011): Ouverture d’ensemble, repli des élites. *Revue française de sociologie*, 55(3), 459–505.
- Bratsberg, B., Markussen, S., Raaum, O., Røed, K., & Røgeberg, O. (2023). Trends in Assortative Mating and Offspring Outcomes. *The Economic Journal*, 133(651), 928–950.
<https://doi.org/10.1093/ej/ueac071>
- Breen, R., & Salazar, L. (2011). Educational assortative mating and earnings inequality in the United States. *American Journal of Sociology*, 117(3), 808–843.
- CEDLAS, & World Bank. (2024). *Socio-Economic Database for Latin America and the Caribbean*.
<https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/sedlac/estadisticas/>
- Chiappori, P.-A., Costa Dias, M., & Meghir, C. (2021). *The measuring of assortativeness in marriage: A comment*.
- Chiappori, P.-A., Costa Dias, M., Meghir, C., & Zhang, H. (2025). *Changes in Marital Sorting: Theory and Evidence from the US*. National Bureau of Economic Research.
- Ciscato, E., & Weber, S. (2020). The role of evolving marital preferences in growing income inequality. *Journal of Population Economics*, 33(1), 307–347.
<https://doi.org/10.1007/s00148-019-00739-4>

- De La Mata, D., Berniell, L., Schargrodsky, E., Álvarez, F., & Alves, G. (2022). *Desigualdades heredadas. El rol de las habilidades, el empleo y la riqueza en las oportunidades de las nuevas generaciones*. Caracas.
- Eika, L., Mogstad, M., & Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6), 2795–2835.
- Esteve, A., Lesthaeghe, R., & López-Gay, A. (2012). The Latin American Cohabitation Boom, 1970–2007. *Population and Development Review*, 38(1), 55–81. <https://doi.org/10.1111/j.1728-4457.2012.00472.x>
- Fernandez, R., Guner, N., & Knowles, J. (2005). Love and money: A theoretical and empirical analysis of household sorting and inequality. *The Quarterly Journal of Economics*, 120(1), 273–344.
- Funes Leal, V. E. (2015). *Emparejamiento selectivo y desigualdad en Argentina* [Tesis, Universidad Nacional de La Plata]. <https://doi.org/10.35537/10915/51560>
- Gabrielli, M. F., & Serio, M. (2017). TESTING ASSORTATIVE MATING: EVIDENCE FROM ARGENTINA. *Revista de Análisis Económico*, 32(2), 109–129. <https://doi.org/10.4067/S0718-88702017000200109>
- Ganguli, I., Hausmann, R., & Viarengo, M. (2014). Marriage, education and assortative mating in Latin America. *Applied Economics Letters*, 21(12), 806–811.
- García, B., & Rojas, O. L. (2004). *Las uniones conyugales en América Latina: Transformaciones en un marco de desigualdad social y de género*. <https://repositorio.cepal.org/bitstreams/46858a47-39f8-4373-a880-21f8e0a018e2/download>
- Gihleb, R., & Lang, K. (2020). *Educational homogamy and assortative mating have not increased*. Emerald Publishing Limited.
- González, F. A. I., & Dip, J. A. (2022). The Impact of Educational Assortative Mating on Income Inequality: Evidence from Argentina. *Journal of Population and Social Studies [JPSS]*, 30, 116–127.
- Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2014). Marry your like: Assortative mating and income inequality. *American Economic Review*, 104(5), 348–353.

- Halliday, T., Mazumder, B., & Wong, A. (2021). Intergenerational mobility in self-reported health status in the US. *Journal of public economics*, 193, 104307.
- Handy, C. (2021). A rank-based measure of educational assortative mating. *Applied Economics Letters*, 28(12), 1031–1035. <https://doi.org/10.1080/13504851.2020.1796907>
- Harmenberg, K. (2014). A note: The effect of assortative mating on income inequality. *Instituto de Estudios Económicos Internacionales-Universidad de Estocolmo*.
- Hryshko, D., Juhn, C., & McCue, K. (2017). Trends in earnings inequality and earnings instability among US couples: How important is assortative matching? *Labour Economics*, 48, 168–182.
- IPUMS. (2020). *Integrated Public Use Microdata Series, International: Version 7.2* (p. <https://www.ipums.org/projects/ipums-international/d020.V7.2>) [Base de Datos]. <https://doi.org/10.18128/D020.V7.2>
- Kremer, M. (1997). How much does sorting increase inequality? *The Quarterly Journal of Economics*, 112(1), 115–139.
- Pereira, R. H. M., Braga, C. K. V., Serra, B., & Nadalin, V. G. (2020). Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras - 2019. *Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras - 2019*, 50–50.
- Quilodrán, J. (2001). *Un siglo de matrimonio en México*. El Colegio de México. https://muse.jhu.edu/pub/320/oa_monograph/book/74397
- Schwartz, C. R., & Mare, R. D. (2005). Trends in educational assortative marriage from 1940 to 2003. *Demography*, 42(4), 621–646.
- Shen, J. (2019). *(Non-) Marital Assortative Mating and the Closing of the Gender Gap in Education*, "Working Paper.
- Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 1013–1024.
- Torche, F. (2010). Educational assortative mating and economic inequality: A comparative analysis of three Latin American countries. *Demography*, 47(2), 481–502.
- Torche, F. (2014). Intergenerational Mobility and Inequality: The Latin American Case. *Annual Review of Sociology*, 40(1), 619–642. <https://doi.org/10.1146/annurev-soc-071811-145521>

Apéndice

Tablas

Tabla A.1 Censos disponibles por país – década.

País	Década 1960	Década 1970	Década 1980	Década 1990	Década 2000	Década 2010
Argentina	-	1970	1980	1991	2001	2010
Bolivia	-	1976	-	1992	2001	2012
Brasil	1960	1970	1980	1991	2000	2010**
Chile	1960*	1970	1982	1992	2002	2017
Colombia	1964*	1973	1985	1993	2005	-
Costa Rica	1963*	1973	1984	-	2000	2011
Dom. Rep.	1960*	-	1981	-	2002	2010
Ecuador	1962*	1974	1982	1990	2001	2010
El Salvador	-	-	-	1992	2007	-
Guatemala	1964	1973	1981	1994	2002	-
Haití	-	1971	1982	-	2003	-
Honduras	1961*	1974	1988	-	2001	-
Jamaica	-	-	1982	1991	2001	-
México	1960*	1970		1990/1995	2000/2005	2010/2015
Nicaragua	-	1971	-	1995	2005	-
Panamá	1960	1970	1980	1990	2000	2010
Paraguay	1962	1972	1982	1992	2002	-
Perú	-	-	-	1993	2007	2017
Puerto Rico	-	1970	1980	1990**	2000**/2005**	2010
Trinidad y Tobago	-	-	1980	1990	2000	2011
Uruguay	1963	1975	1985	1996	2006	2011**
Venezuela	-	1971	1981	1990	2001	-
Total	6	17	17	17	21	11

Nota: la tabla presenta el año de censo correspondiente a cada década y país. Los años marcados con * son censos en los cuales no es posible vincular a los habitantes del hogar. Aquellos marcados con ** no tienen información de los años de escolaridad.

Fuente: elaboración propia en base a IPUMS (2020).

Tabla A2 – Estadísticas descriptivas período completo

Panel A. En pareja												
Medias muestrales	1960		1970		1980		1990		2000		2010	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Share (%)	64.92	64.06	63.86	65.63	61.02	63.07	60.13	61.31	58.08	59.07	52.88	51.42
Edad	34.16	35.13	34.21	35.10	33.95	34.77	34.20	34.99	34.73	35.48	35.11	35.71
Años de escolaridad	3.08	3.53	3.84	4.51	5.89	6.42	6.72	7.24	7.69	7.95	10.25	10.14
PRII (%)	78.14	76.02	68.32	63.29	46.43	42.01	37.37	32.45	31.15	27.69	13.42	12.10
PRIC (%)	18.30	18.84	24.19	26.58	37.05	39.76	40.73	44.26	36.92	40.27	34.21	37.02
SEC (%)	3.14	3.49	6.16	7.25	13.77	13.57	18.24	17.88	26.38	25.81	37.27	38.06
UNIC (%)	0.42	1.65	1.33	2.88	2.75	4.66	3.67	5.41	5.54	6.23	15.11	12.82
Obs.	1424584	1421867	2802242	2778991	3613742	3660534	3954365	3853054	5873289	5636849	3119801	2920138
Panel B. Solteros												
Medias muestrales	1960		1970		1980		1990		2000		2010	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Share (%)	35.08	35.94	36.14	34.37	38.98	36.93	39.87	38.69	41.92	40.93	47.12	48.58
Edad	33.70	32.30	33.64	32.26	33.06	31.80	33.14	31.99	33.55	32.48	33.64	32.92
Años de escolaridad	3.62	3.69	4.26	4.63	6.49	6.53	7.56	7.58	8.65	8.35	10.93	10.38
PRII (%)	72.30	74.13	63.66	61.88	41.26	40.77	31.20	30.12	25.53	26.01	11.42	12.47
PRIC (%)	21.99	20.30	26.70	27.02	37.25	39.48	38.72	42.45	33.04	37.13	28.18	32.42
SEC (%)	4.85	3.77	7.95	8.15	17.50	15.15	24.32	21.41	32.75	29.25	41.37	40.77
UNIC (%)	0.87	1.80	1.69	2.95	3.99	4.59	5.75	6.02	8.69	7.61	19.03	14.33
Obs.	528257	519202	1190928	1150934	1627503	1457335	2012951	1818742	3377285	3021956	2198090	1966669

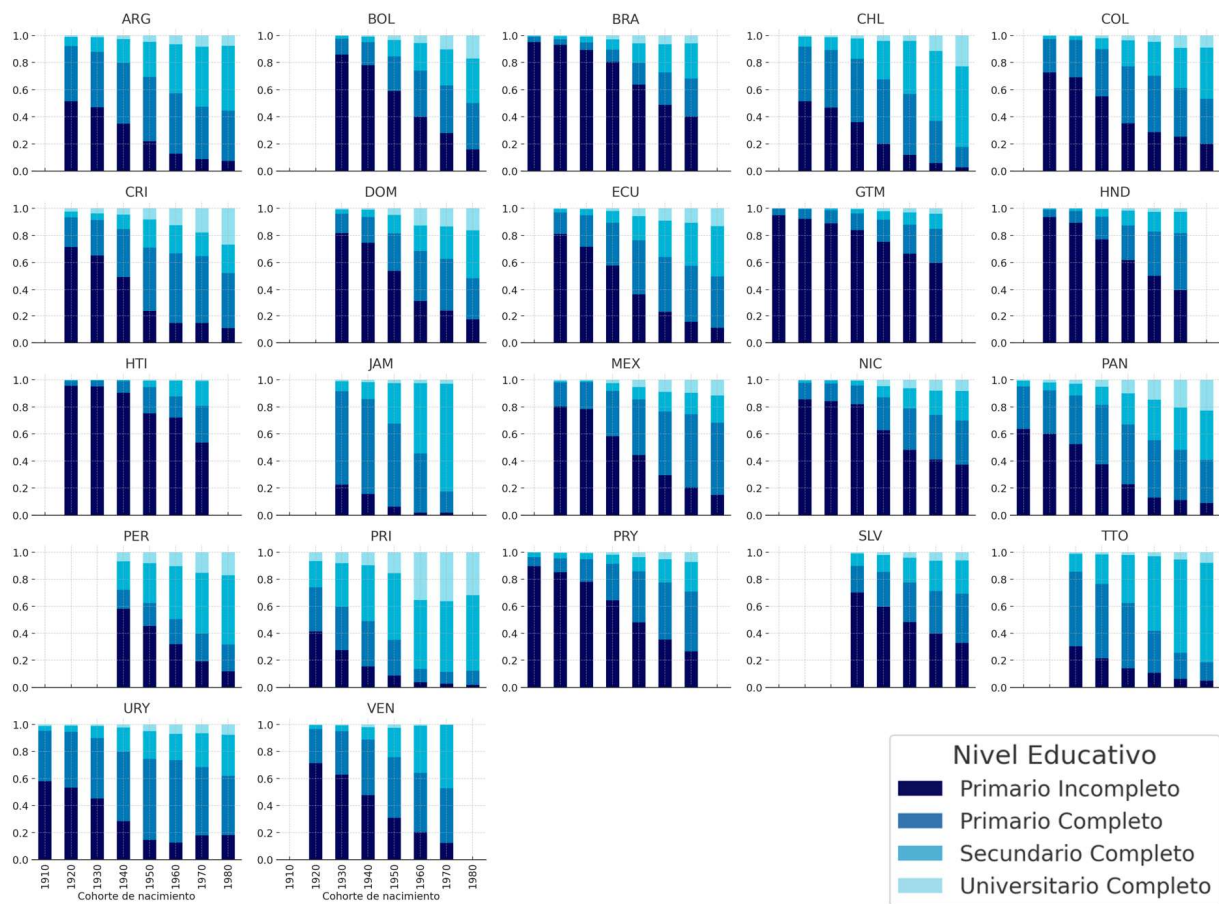
Nota: La tabla presenta el promedio regional de medias muestrales de edad y educación, así como también el total de observaciones para cada década censo. Se divide a la muestra entre individuos solteros y aquellos en pareja. Es importante destacar que los países incluidos en cada década censo pueden variar dependiendo de la disponibilidad de datos para cada década y país.

Fuente: IPUMS-I (2020)

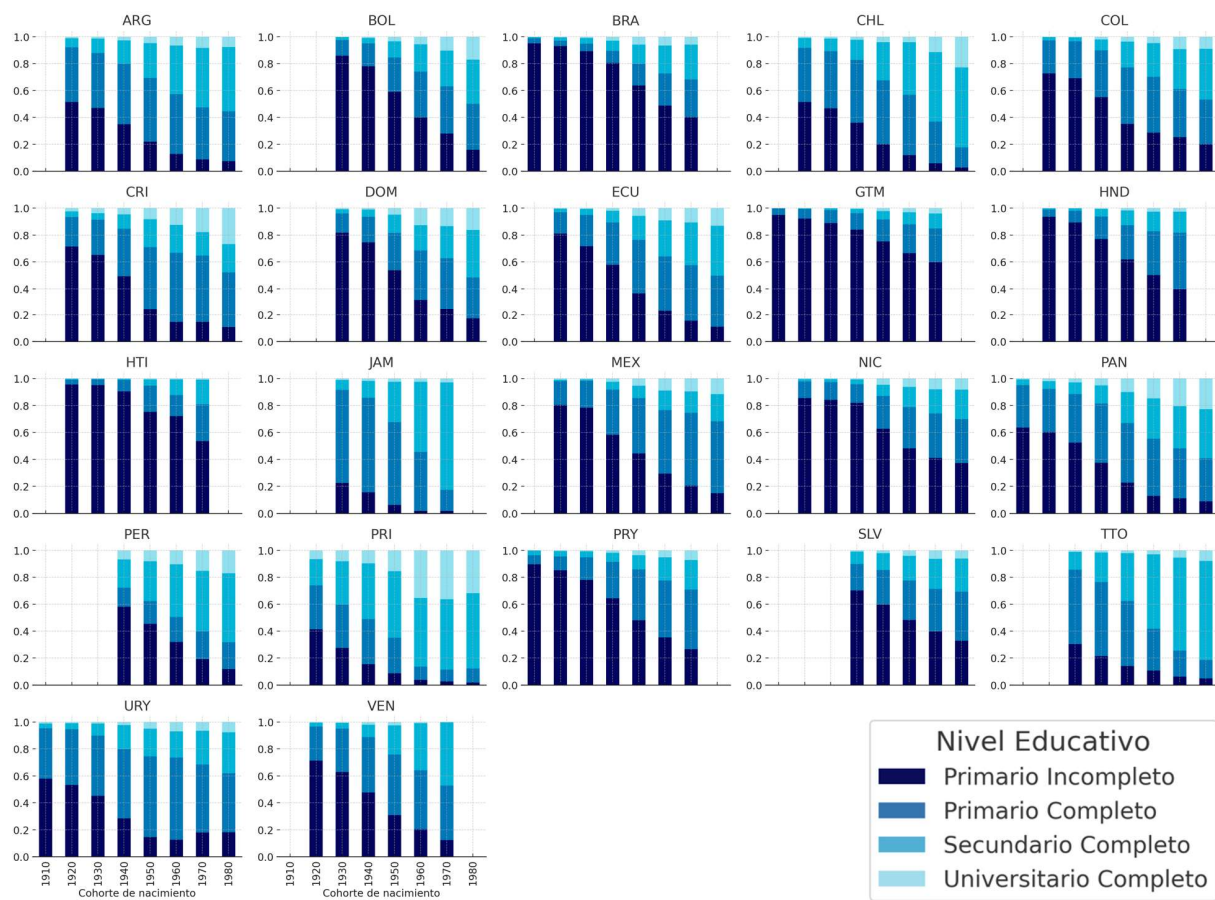
Figuras

Figura A1 - Distribución educativa de hombres por país y cohorte de nacimiento.

Panel A. Hombres



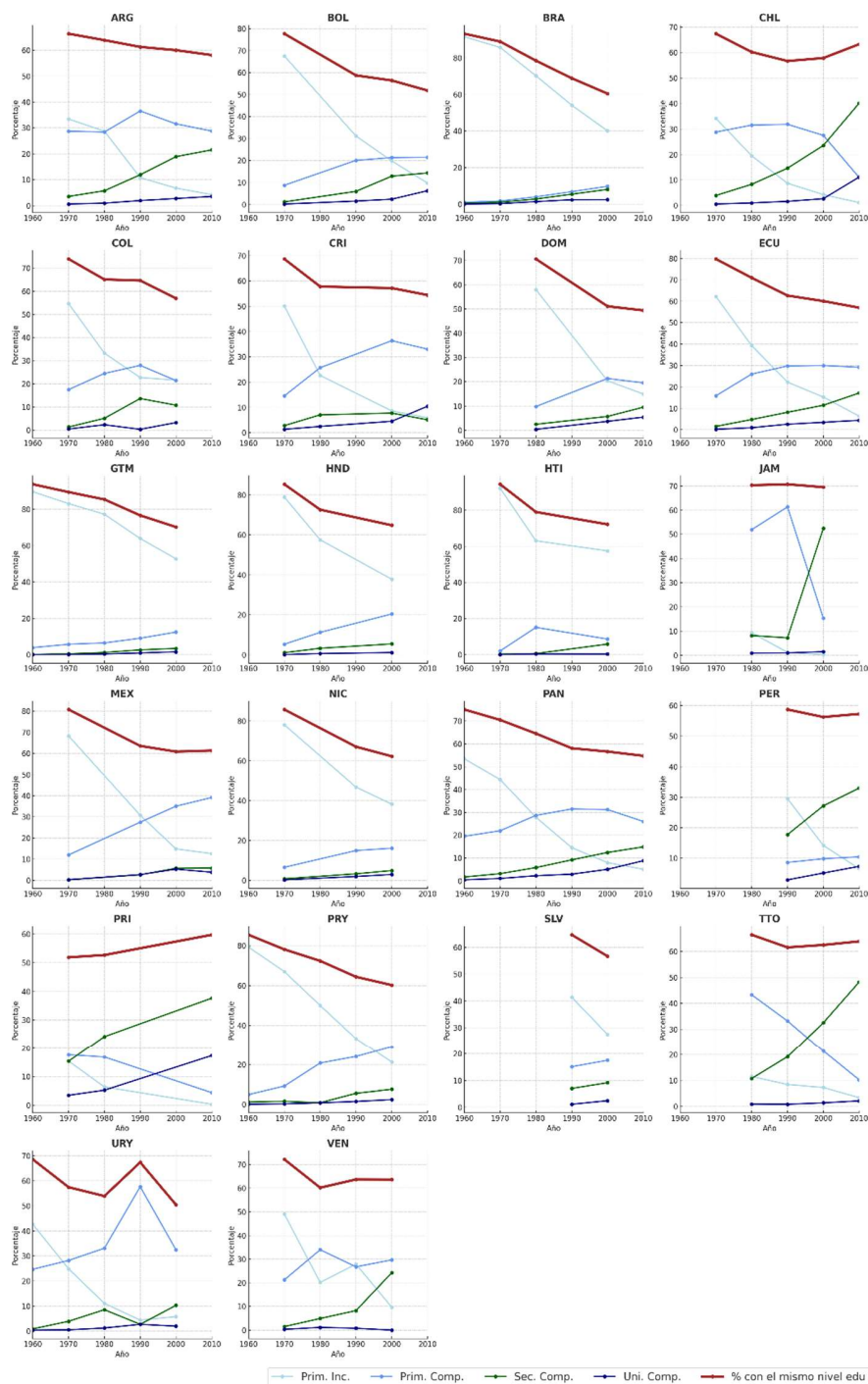
Panel B. Mujeres



Nota: cada panel muestra el porcentaje de hombres según máximo nivel educativo alcanzado, por país por década de censo. En el primer panel, se incluyen hombres de entre 25 y 45 años; en el segundo, mujeres de ese mismo rango etario.

Fuente: IPUMS-I (2020)

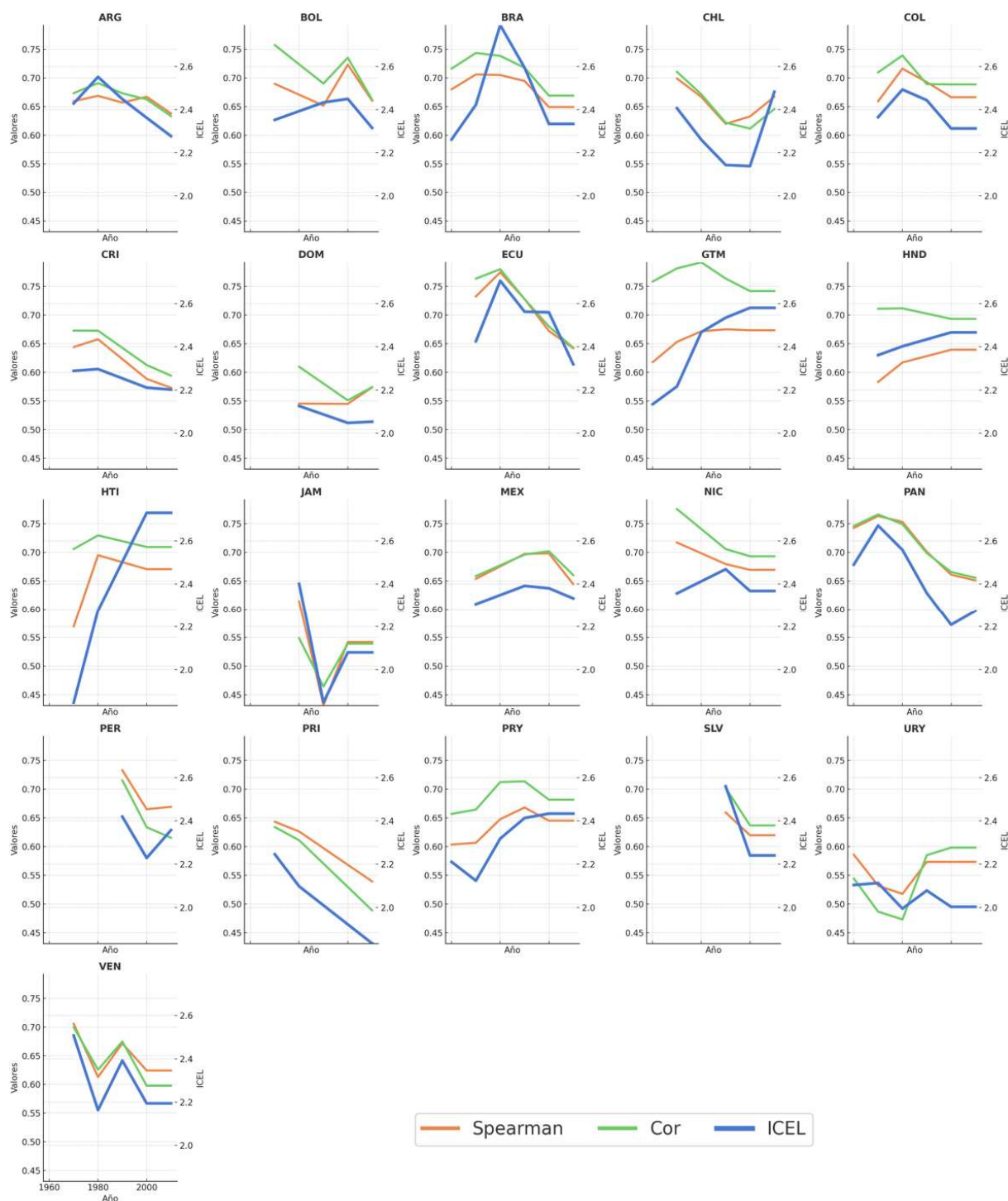
Figura A2. Porcentaje de parejas por nivel educativo – por país y por década de censo



Nota: esta figura presenta el porcentaje de parejas con el mismo nivel educativo, desglosado por nivel educativo y para el total de la población. Los países incluidos pueden variar según la disponibilidad de censos.

Fuente: IPUMS-I (2020).

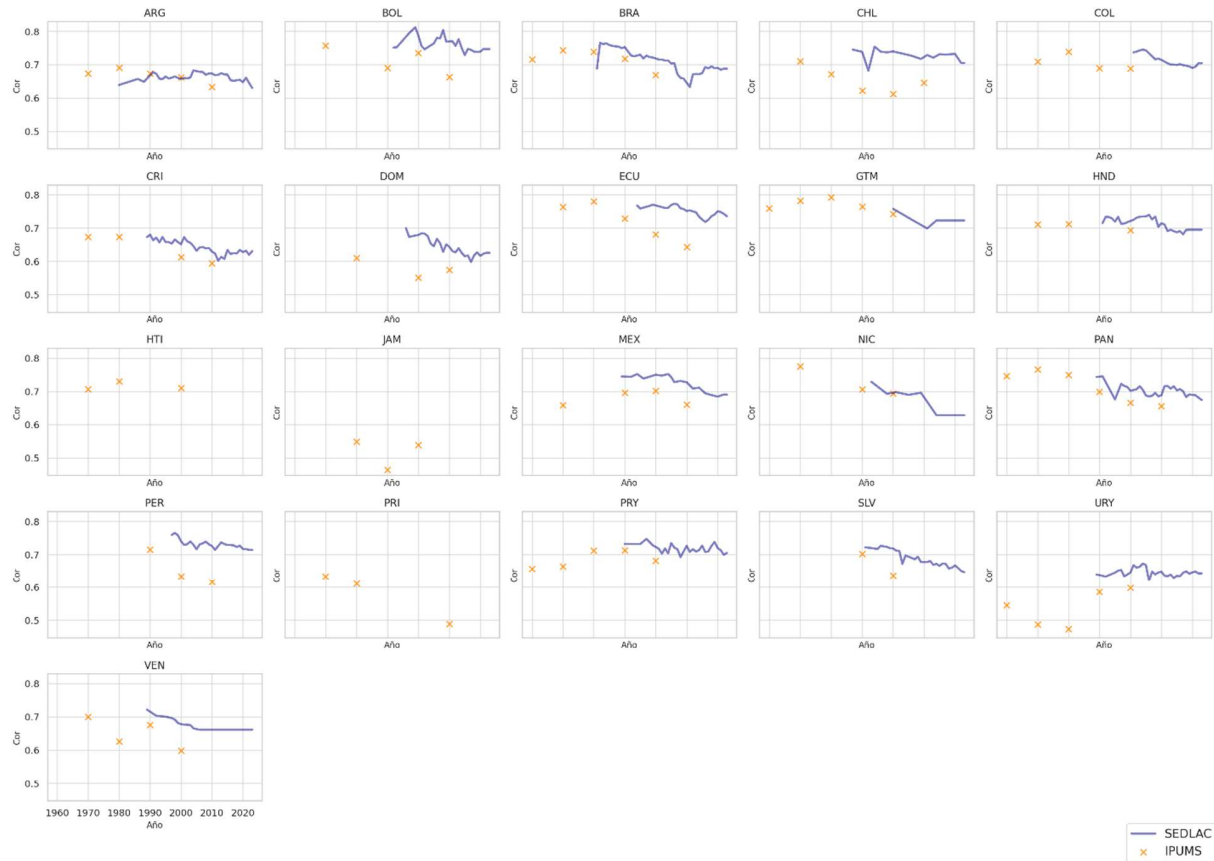
Figura A3. Evolución del ES según indicadores alternativos:



Nota: Cada panel del gráfico muestra la evolución de los indicadores de correlación de Spearman, Pearson y del ICEL para cada uno de los años con información disponible.

Fuente: Elaboración propia en base a IPUMS (2020).

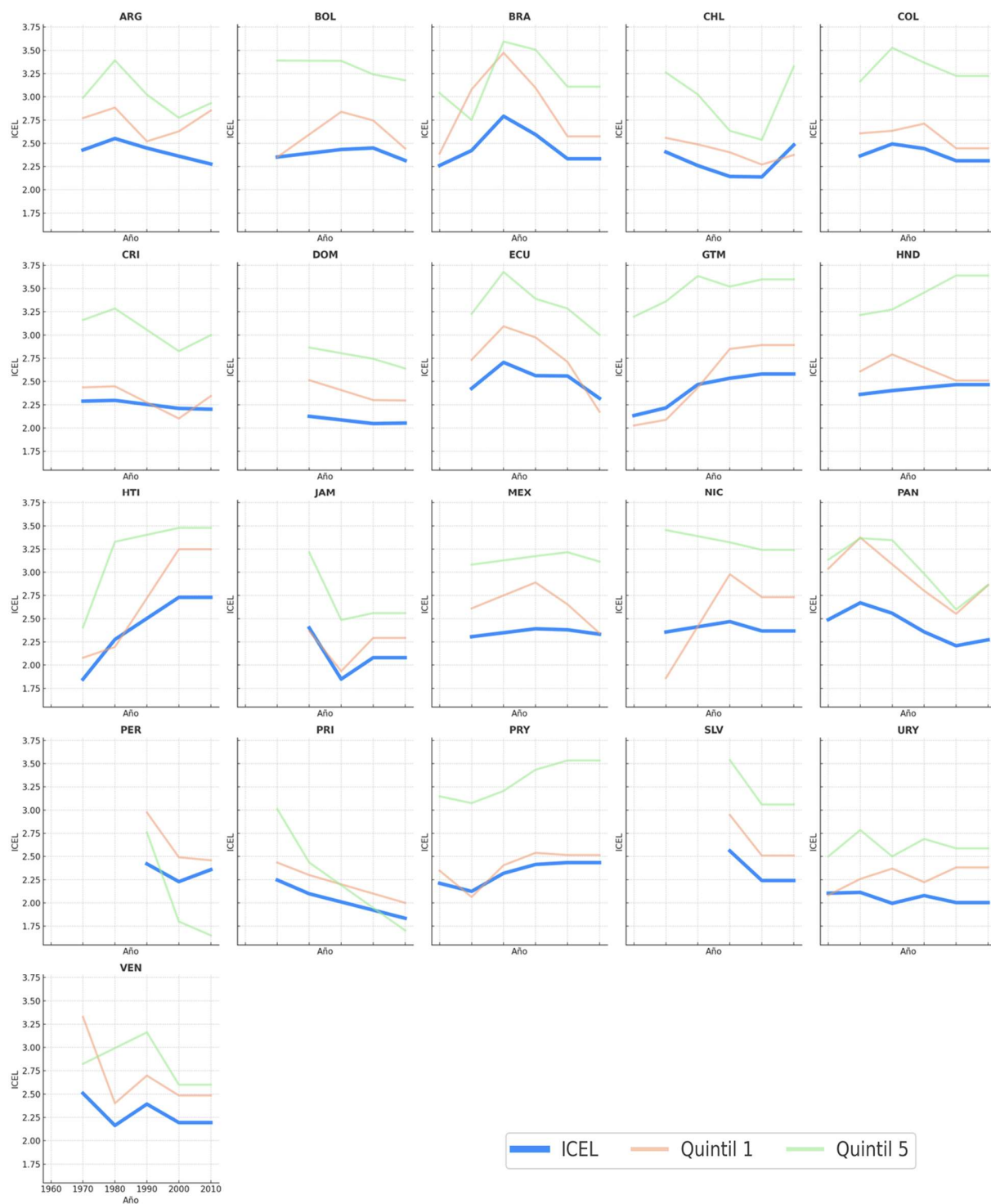
Figura A4. Coeficiente de correlación de Pearson según fuentes alternativas



Nota: este gráfico presenta los valores del Coeficiente de Correlación Lineal de Pearson utilizando estimaciones provenientes de dos fuentes de datos alternativas. Los puntos representan los valores provenientes de las estimaciones de IPUMS. Las líneas, los valores de SEDLAC (CEDLAS y Banco Mundial).

Fuente: Elaboración propia en base IPUMS (2020) y a SEDLAC (2024).

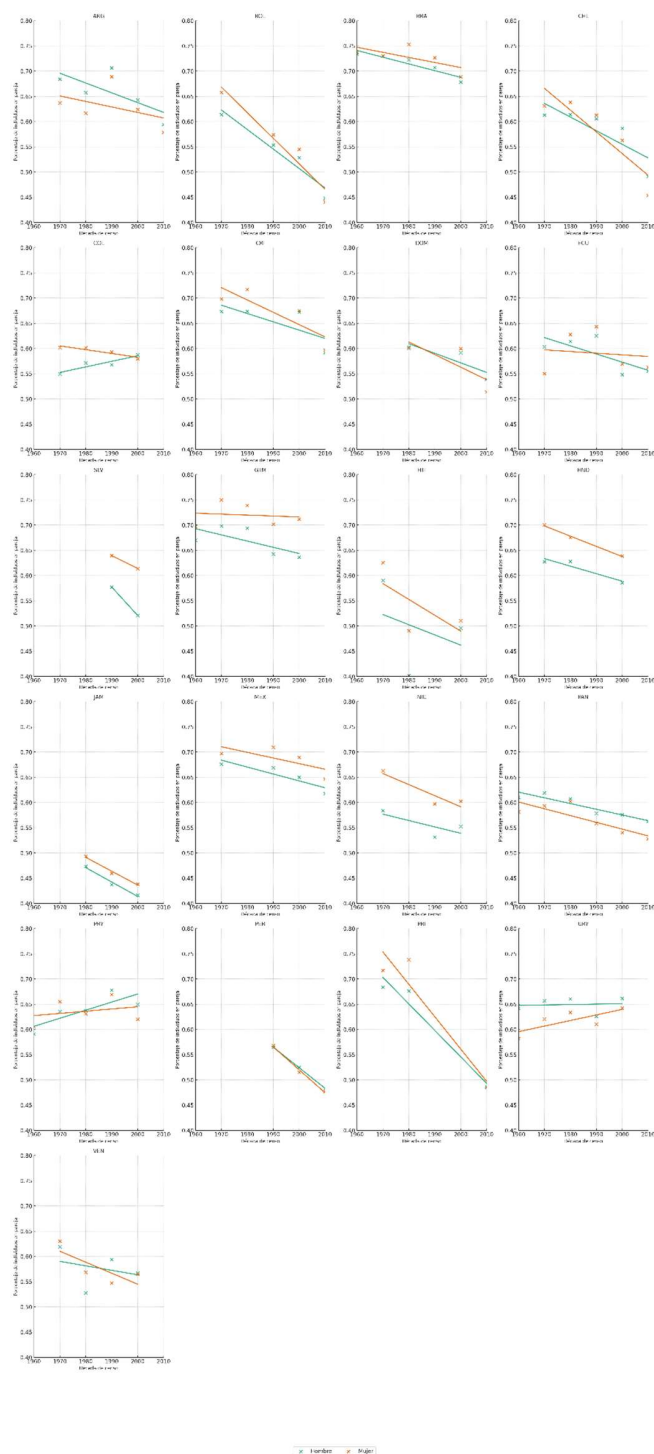
Figura A5. Evolución del ICEL



Nota: Cada panel del gráfico muestra la evolución de ICEL agregado, así como el ICEL Q1 y Q5 para cada uno de los años con información disponible.

Fuente: Elaboración propia en base a IPUMS (2020).

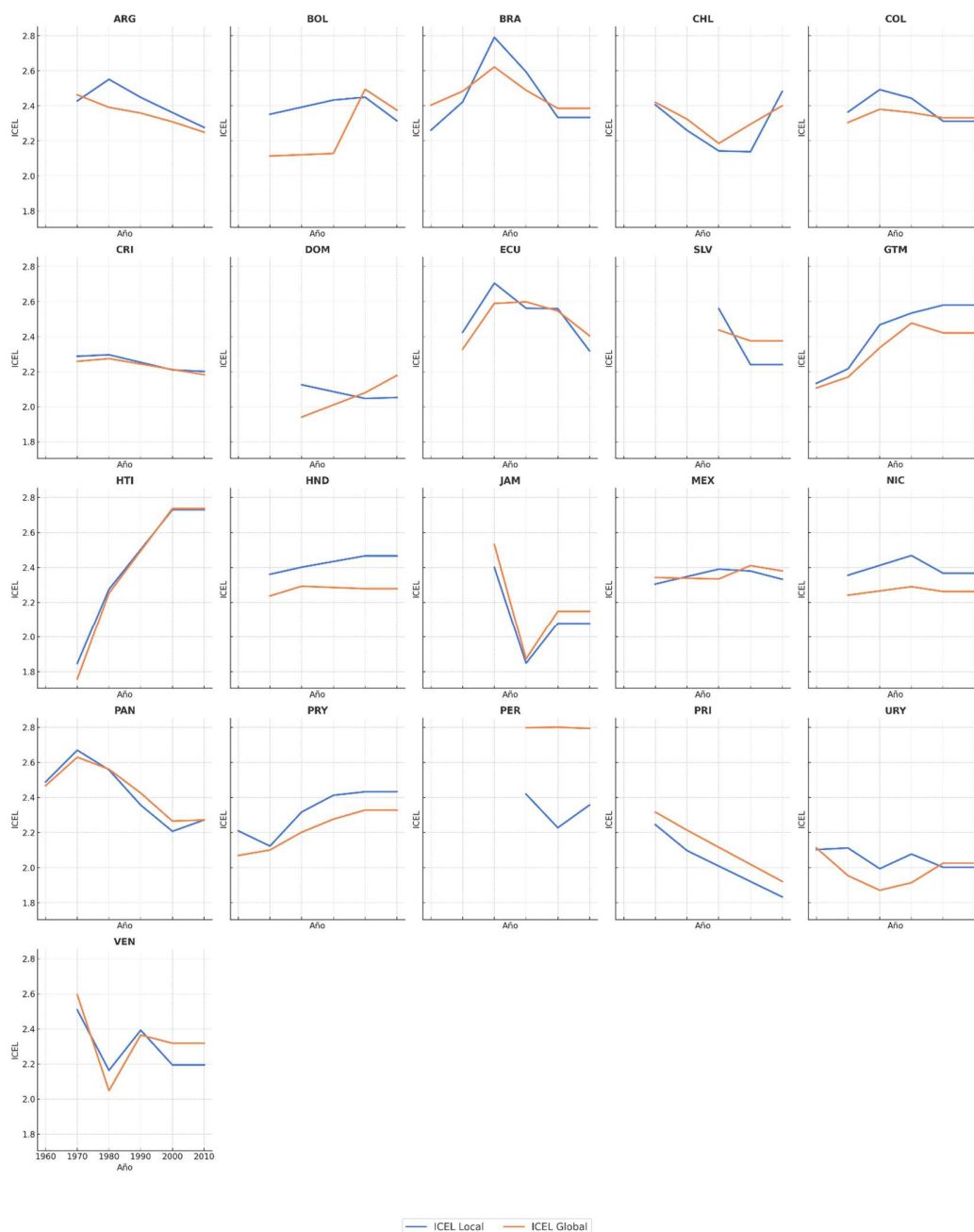
Figura A6. Porcentaje de individuos en pareja, por país y censo.



Nota: la figura presenta el share de individuos de entre 25 y 45 años en pareja en relación al total de la población.

Fuente: elaboración propia en base a IPUMS (2020)

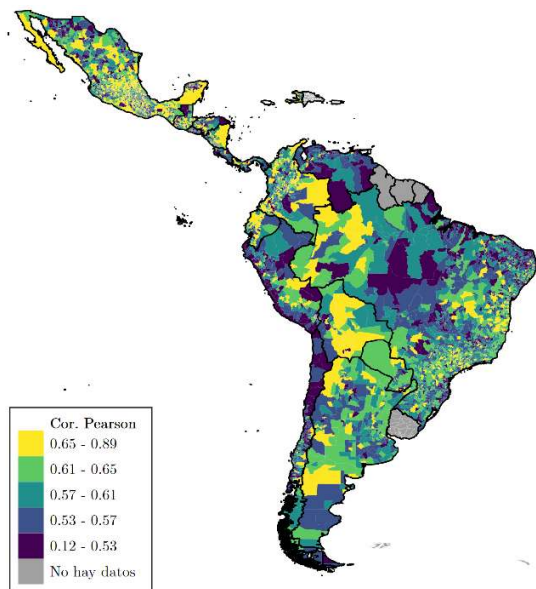
Figura A7. Evolución ICEG vs ICEL



Nota: la figura presenta la evolución del ICEL y del ICEG agregado por década censo.

Fuente: elaboración propia en base a IPUMS (2020)

Figura A8. Coeficientes de correlación lineal educativa por Geolevel 2 –LAC, Década 2000



Nota: el gráfico presenta los valores que toma el coeficiente de correlación de Pearson de años educativos de hombre y mujeres en pareja, por unidad subnacional de segundo orden en la década de los 2000.

Fuente: elaboración propia en base a IPUMS (2020)

Apéndice I - Identificación de parejas

Para poder calcular las medidas de emparejamiento selectivo es necesario observar la educación de los ambos miembros de una pareja. Para eso, es necesario identificar a los miembros que conforman cada pareja.

Si bien esta fuente de datos cuenta con una variable denominada “SPLOC”, que permite identificar diferentes parejas dentro de un hogar, esta no se encuentra disponible en algunos de los censos, motivo por el que no se utilizará. En su lugar, se utilizará la variable “RELATE”, la cual identifica al jefe de hogar y todas las relaciones vinculadas con este. Si bien este criterio no permite establecer conexiones de parejas entre el resto de los miembros del hogar diferentes al jefe, creemos que es el más adecuado para un mejor estudio de las tendencias, ya que esta variable se encuentra disponible en todos los censos utilizados en el análisis.

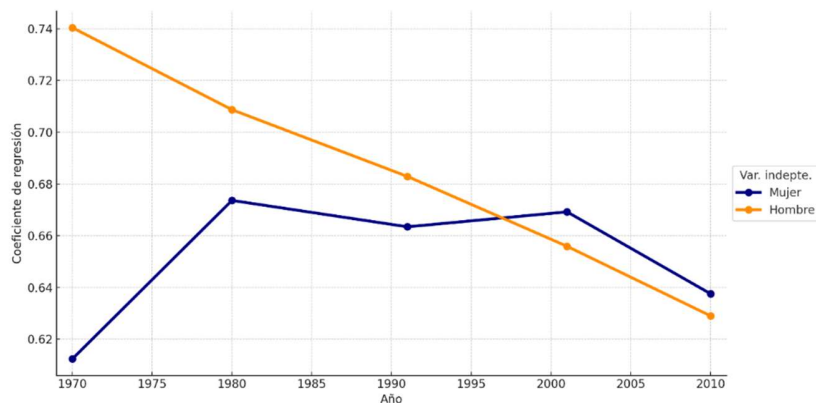
Apéndice II – Regresión y regresión inversa

El modelo que se estima consiste en una regresión en la cual la variable dependiente son los años de educación de la esposa ($Yedu_{esposa}$) y la independiente, los años de educación del esposo ($Yedu_{esposo}$): $Yedu_{esposa} = \alpha + \beta Xedu_{esposo} + u$.

Por definición, el coeficiente de regresión es: $\beta = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_y}{\sigma_x} \rho$. Donde σ es el desvío estándar de cada una de las distribuciones y ρ es el coeficiente de correlación de Pearson. Esto implica que un aumento en el coeficiente de regresión puede reflejar un aumento en el coeficiente de correlación de Pearson o un aumento en la varianza del logro educativo de las mujeres en relación con la varianza de la educación de los hombres (si se considera la especificación anterior).

Para mostrar la invalidez de este indicador para medir la evolución del ES, a continuación, se estudiará la evolución de los coeficientes de regresión a lo largo del tiempo para Argentina, considerando alternativamente como variables independientes la educación de esposos y de esposas, respectivamente. La Figura A2.1 muestra las tendencias que surgirían si se considerara la educación de la esposa, además de la típicamente computada educación del hombre, como variable independiente.

Figura A2.1- Estimación de emparejamiento selectivo a través de MCO cambiando la variable independiente en Argentina.



Nota: el gráfico presenta los resultados de una regresión en la cual la variable dependiente es la educación del hombre (mujer) y la independiente la de la mujer (hombre), por año de censo disponible. Se consideran parejas en las cuales al menos uno de los cónyuges tiene entre 25 y 45 años.

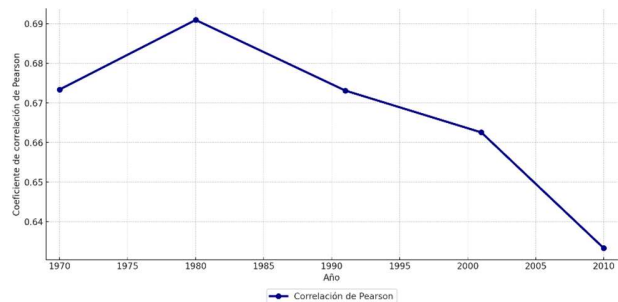
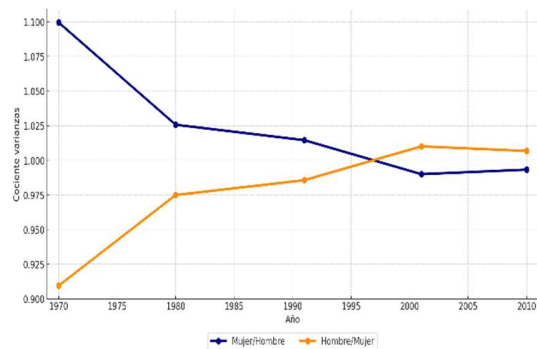
Fuente: elaboración propia en base a IPUMS (2020)

Mientras que el coeficiente de la educación de la esposa disminuye con el tiempo, el coeficiente de del esposo aumenta durante el mismo período. El motivo que explica estas diferencias se vincula con que la varianza en los años de educación entre los hombres cayó sustancialmente durante el período en consideración, en comparación con la de los años de educación entre las mujeres.

Figura A2.2. Evolución de los componentes del coeficiente de regresión en Argentina.

Panel A. Cociente de varianzas

Panel B. Coeficiente de correlación de Pearson



Nota: el panel A presenta el cociente de desvíos estándar entre la distribución educativa de mujeres y hombres, alternando el numerador y el denominador. El panel B presenta la evolución del cociente de correlación lineal de Pearson.

Fuente: elaboración propia en base a IPUMS (2020)

La comparación de los resultados de las Figuras A2.1 y A2.2 muestra cómo los cambios en las distribuciones marginales de la educación, si se ignoran, pueden conducir a conclusiones injustificadas sobre la evolución del emparejamiento educativo selectivo.

Apéndice III. Indicadores de segunda generación

Ratio de probabilidad y modelos log-lineales

El ratio de probabilidad es una medida utilizada para evaluar el emparejamiento selectivo en términos de la probabilidad de que dos individuos con características similares formen una pareja, en comparación con aquellos que poseen características diferentes. Esta métrica permite identificar si existe una tendencia a que personas con niveles educativos semejantes se emparejen con mayor frecuencia de lo esperado por azar. El ratio de probabilidad (Odds Ratio, OR) se define como el cociente entre:

- La probabilidad de formar una pareja homogámica (es decir, una pareja en la que ambos miembros tienen el mismo nivel educativo).
- La probabilidad de formar una pareja no homogámica (una pareja en la que los miembros tienen niveles educativos diferentes).

El ratio de probabilidad se puede expresar mediante la siguiente fórmula:

$$IO(a, b, c, d) = \ln \left(\frac{a \cdot d}{b \cdot c} \right)$$

El Odds Ratio es ampliamente utilizado en la literatura demográfica, ya que puede derivarse directamente del enfoque log-lineal (véanse, por ejemplo, Mare (2001), Mare y Schwartz (2005), Bouchet-Valat (2014)). En economía, ha sido empleado por Siow (2015) y por trabajos como Chiappori et al. (2017), Ciscato y Weber (2020), y Chiappori, Costa-Dias, Crossman y Meghir (2020), entre otros.

El modelo log lineal es ampliamente utilizado en la sociología para medir el Emparejamiento selectivo. La hipótesis nula del modelo log-lineal es la independencia estadística completa (es decir, emparejamiento aleatorio). En particular, sea i el nivel educativo del esposo y j el nivel educativo de la esposa, y sea μ_{ijkl} el número de parejas donde el esposo tiene educación i y la esposa tiene educación j . Entonces, la independencia estadística significa que:

$$\mu_{ij} = A \cdot \alpha_i \cdot \beta_j$$

Donde A, α_i , y β_j son parámetros a ajustar. Usualmente, los investigadores toman logaritmos de la ecuación anterior, de manera que:

$$\log(\pi_{ij}) = A' + \alpha'_i + \beta'_j$$

Donde nuevamente, A', α'_i , and β'_j son parámetros a ajustar, pero ahora pueden ser ajustados. Un término residual adicional añadido a la Ecuación 1.4 captura cuán assortativos son los emparejamientos, netos de lo que se predice por emparejamiento aleatorio. Por ejemplo, para medir las tendencias agregadas, Schwartz y Mare (2005) añaden una variable dummy igual a 1 si la educación del esposo y la de la esposa son

iguales y cero en caso contrario (el "modelo de homogamia", ecuación (2)). Para aislar partes particulares de la distribución educativa que están impulsando las tendencias agregadas, estiman un "modelo de cruces", donde añaden varias variables dummy que estiman la dificultad relativa de cruzar cada límite educativo (Schwartz y Mare (2005), ecuación (3)).

Propuesta de Eika:

La propuesta de Eika, et al. (2019) sugiere medir el Emparejamiento Selectivo educativo a través de la comparación de tablas de contingencia. Plantea que, a partir de estas tablas, es posible medir el emparejamiento marital entre los niveles de educación del esposo e_o y la esposa e_a como la probabilidad observada de que un esposo con nivel de educación e_o esté casado con una esposa con nivel de educación e_a , en relación con la probabilidad que surgiría bajo emparejamiento aleatorio. Formalmente, se define al indicador $s(e_o, e_a)$ como:

$$s(e_o, e_a) = \frac{P(E_o = e_o, E_a = e_a)}{P(E_o = e_o)P(E_a = e_a)},$$

donde $E_o(E_a)$ denota el nivel educativo del esposo (esposa). El emparejamiento selectivo positivo se produce entonces cuando el indicador $s(e_o, e_a)$ toma valores mayores a 1, lo cual significa que hombres y mujeres con niveles de educación e_o y e_a se casan con más frecuencia de lo que se esperaría bajo un patrón de emparejamiento aleatorio en términos de educación. Para obtener una medida del emparejamiento selectivo educativo agregado, se calcula el promedio ponderado de los parámetros de emparejamiento a lo largo de la diagonal principal.

Ejercicio de simulación

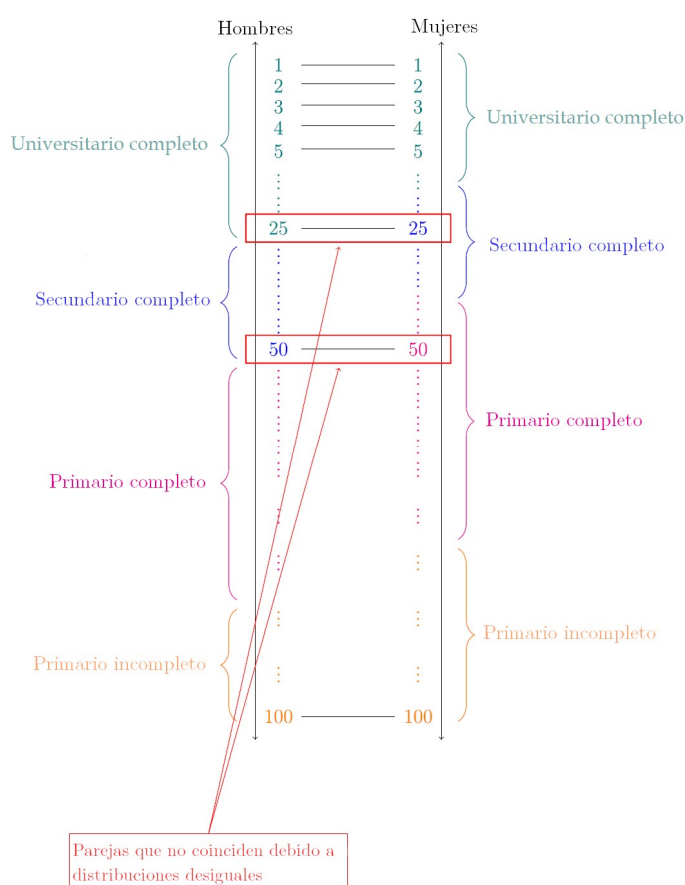
Una manera de evaluar la capacidad que tiene el indicador propuesto por Eika et al., (2019) para controlar completamente los cambios en las distribuciones educativas de hombres y mujeres es mediante la estimación del mismo en diferentes escenarios, en los cuales se asume un emparejamiento selectivo perfecto, pero se varían los supuestos en relación a las distribuciones educativas. En esta sección, se presentará un ejercicio de simulación basado en el trabajo de Shen (2019), adaptado a los datos de Argentina¹⁸, y se incluirán nuevos escenarios, que incorporan las críticas de Bratsberg et al., (2018). A continuación, se evaluará el indicador mediante su funcionamiento en cuatro escenarios alternativos: el observado y tres simulados.

En la primera simulación, se tomará la distribución educativa de hombres y mujeres, dividida en los cuatro niveles presentados en la sección anterior, para cada uno de los años bajo análisis, y se asumirá un emparejamiento perfecto. En cada año t , un número igual de hombres y mujeres se rankean por separado según su nivel educativo, de modo que hay un

¹⁸ Si bien se podría haber elegido cualquiera de los 22 países disponibles en los censos, la información para Argentina presenta la ventaja de que se encuentra disponible para cinco décadas a lo largo del tiempo, lo cual permite estudiar las tendencias sobre la evolución del fenómeno. Además, dicho país, cuenta con un número de observaciones considerable (15 millones en total) pero no excesivo, que permiten realizar los ejercicios de simulación sin demandar grandes costos computacionales (como sí podría suceder en los casos de México o Brasil, por ejemplo).

hombre “top” y una mujer “top”. Cada año, todos forman una pareja y la pareja siempre será perfecta en términos de emparejamiento selectivo: el hombre “top” y la mujer “top” siempre coinciden, al igual que el hombre y la mujer en segundo lugar, y así sucesivamente. Esto es, las personas siempre se emparejarán con alguien que ocupe la misma posición en el ranking de educación correspondiente a su género, sin importar si dicha persona tenga o no el mismo nivel educativo. En la figura A.3.1 se presenta la representación visual de dicho ordenamiento.

Figura A.3.1. Mecanismo de emparejamiento



Fuente: elaboración propia en base a Shen (2019)

Si bien se supone un emparejamiento perfecto, en la primera simulación se permitirá que los tamaños de las categorías educativas difieran según el género y cambien en cada década de acuerdo con las distribuciones educativas observadas en la realidad.

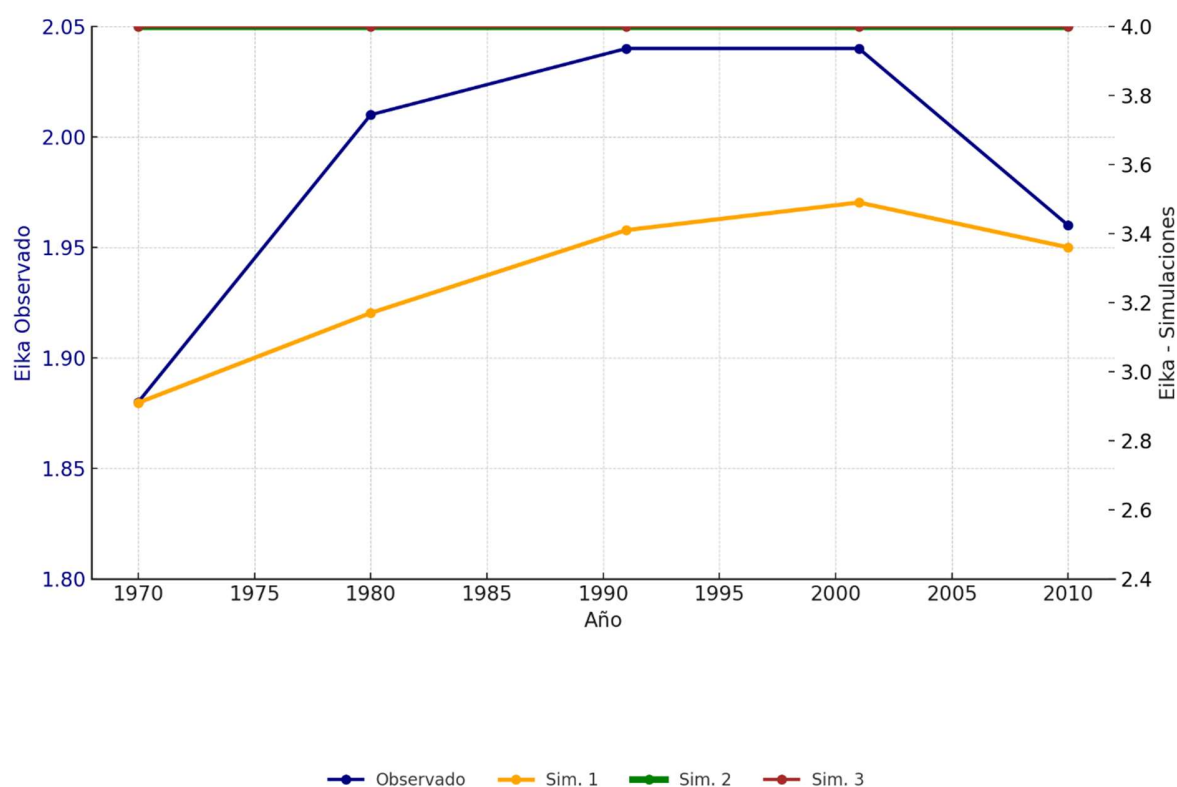
En la segunda simulación, se supondrá un emparejamiento selectivo perfecto y se permitirá que los tamaños de las categorías educativas varíen en el tiempo, pero se asumirá que no existen brechas de género en términos educativos.

En la tercera simulación, además de suponer que existe un Emparejamiento Selectivo perfecto y que no existen brechas de género, se mantendrá constante el tamaño relativo de

dos de las cuatro categorías educativas en cada una de las distribuciones a lo largo del tiempo. En particular, se supondrá que los individuos con universitario completo siempre representan una proporción igual al 5% del total, mientras que aquellos con primario completo representan un 45%. A su vez, la evolución del grupo de individuos con primario incompleto se mantendrá en su estado observado en la realidad, mientras que el porcentaje de individuos con secundario completo responderá a los cambios en el último grupo.

Se evaluará el desempeño del indicador en los cuatro escenarios. Dado que el emparejamiento, por construcción, nunca cambia en los tres escenarios simulados, la métrica debería permanecer constante, incluso cuando cambien las distribuciones de educación de hombres y mujeres. En la Figura A.3.2, se presentan los resultados obtenidos del ejercicio de simulación.

Figura A.3.2. Tendencia del indicador Eika por categorías – Resultados de las simulaciones



Notas: la figura presenta las estimaciones del indicador de Eika para cuatro escenarios alternativos: el observado en la realidad (eje izquierdo) y tres simulados (eje derecho). Es decir, para cada categoría educativa, se calculan las probabilidades (en relación a la coincidencia aleatoria) de observar una pareja en la que ambos miembros tienen ese nivel educativo y se calcula un promedio ponderado de estas probabilidades, donde los ponderadores son la proporción de parejas de educación similar que caen en esa celda. Los datos corresponden a los censos de Argentina.

Fuente: IPUMS-I (2020)

Los resultados de la primera simulación (representados por la línea amarilla) exhiben un aumento en el nivel de emparejamiento selectivo entre los años 1970 y 2010, a pesar de que,

se esperaría que este indicador se mantuviera constante a lo largo de todo el período. De manera llamativa, estos resultados muestran tendencias similares a las observadas en el escenario real.

Por otro lado, en las representaciones en verde y rojo se exhiben los valores del indicador cuando se eliminan las brechas de género a lo largo del tiempo. En estos casos, el indicador funciona de manera correcta, dado que se mantiene constante a lo largo de las décadas y alcanza su nivel máximo¹⁹. Este ejemplo proporciona evidencia de que, si bien el indicador propuesto por Eika logra controlar una parte de las variaciones en la distribución, no logra abordar las variaciones en las brechas de género (Shen, 2019).

Esta limitación se puede explicar considerando el rango de valores que el indicador puede tomar. Este indicador adquiere valores inferiores a 1 cuando no existe emparejamiento selectivo positivo y alcanza un valor igual a 4 cuando el patrón de emparejamiento es perfecto, y todos los valores de la tabla de contingencia se encuentran únicamente en la diagonal principal. Es posible, entonces, que exista un emparejamiento perfecto en términos educativos, pero que el indicador no alcance su valor máximo debido a la existencia de brechas educativas entre hombres y mujeres. En consecuencia, a medida que estas brechas se reduzcan, el valor del indicador aumentará, a pesar de que los patrones de emparejamiento se mantengan constantes. La explicación proporcionada por Shen para este fenómeno está alineada con este razonamiento, ya que sugiere que a medida que las distribuciones educativas de hombres y mujeres evolucionaron y se volvieron más similares, la probabilidad de encontrar un cónyuge con el mismo nivel educativo aumentó. Por esta razón, el indicador resultante de la primera simulación muestra una tendencia similar a la del mundo real y refleja variaciones derivadas de cambios en las brechas educativas.

Para entender la intuición, se presenta el siguiente ejemplo, representado en la Tabla A.3.1. Considere una población de hombres y mujeres cuyas distribuciones educativas se dividen en dos niveles (alta y baja) y son observadas en tres períodos. En el primer período, los hombres se distribuyen de igual manera entre el nivel bajo y alto de educación (50% en cada nivel), mientras que el 70% de las mujeres tiene educación baja y el 30% restante, alta. En ese caso, si existiera emparejamiento perfecto, el 30% de mujeres con educación alta se emparejaría con el 30% de los hombres con mayor educación. Luego, quedaría un 20% de hombres educados que se emparejaría con mujeres de nivel educativo bajo, dado que no encontrarían parejas con el mismo nivel. Por último, dado que el 20% de mujeres de educación baja se emparejó con el nivel más alto, quedaría un 50% de mujeres que se emparejaría con el 50% de hombres restante.

El problema del indicador propuesto radica en que la matriz de contrafactuales utilizada para computarlo tiene en consideración el total de parejas que podrían emparejarse dadas las distribuciones educativas de cada género, pero no considera que, si hay brechas de género, habrá un porcentaje de individuos (en este caso mujeres) que se emparejará con el residual de individuos del otro género (hombres) que no se emparejaron con personas de su mismo

¹⁹ El valor máximo que puede tomar el indicador de Eika es igual al total de categorías educativas en las cuales se divide la distribución educativa. En este caso, si las distribuciones de hombres y mujeres fueran iguales, la matriz aleatoria debería contener valores iguales a $(0.25 \times 0.25 = 0.0625)$, y si el emparejamiento fuera perfecto, las celdas de la diagonal principal deberían ser todas iguales a $(0.25/0.0625 = 4)$

nivel educativo, dado que los géneros no tienen las mismas proporciones. De esa manera, si bien el emparejamiento selectivo es perfecto, el indicador es menor a 2.

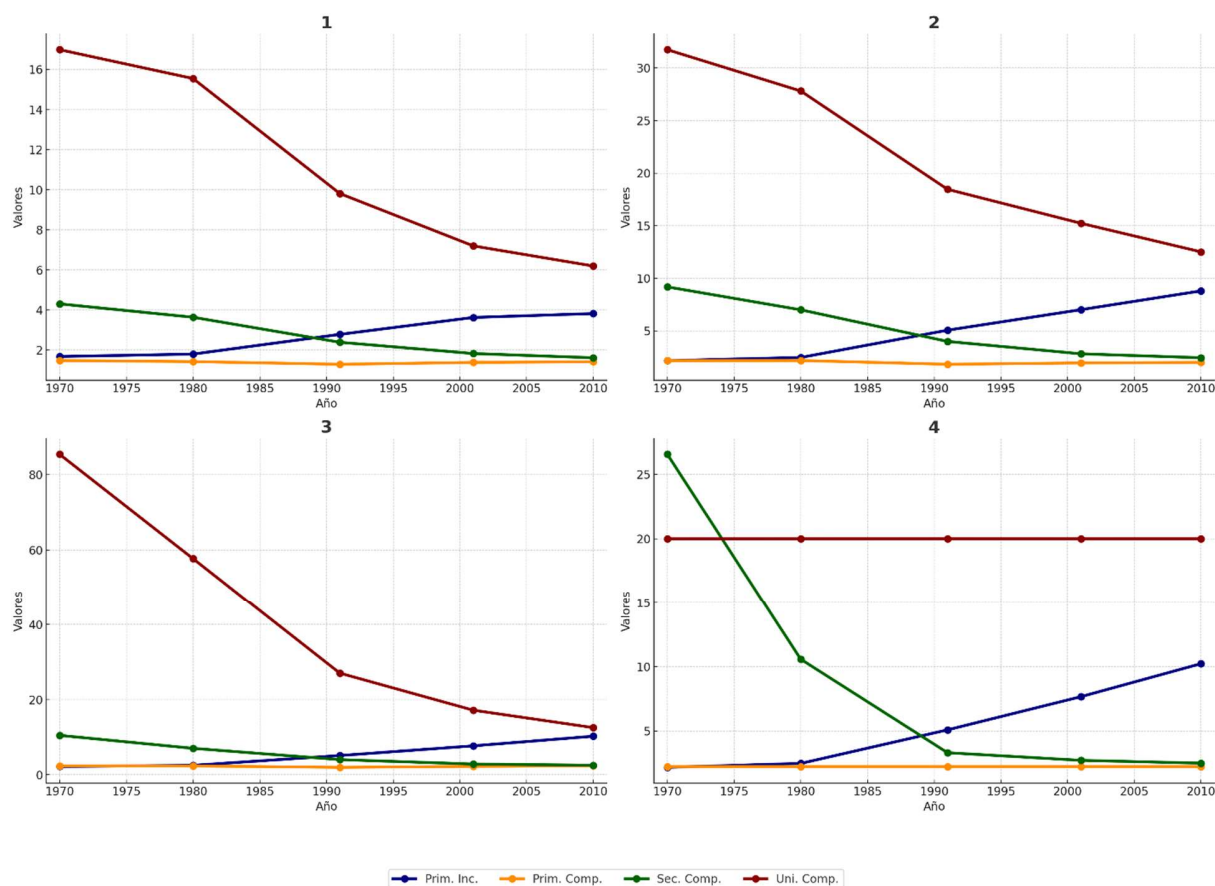
Por otra parte, en el segundo período, el porcentaje de mujeres educadas aumenta en 10 puntos porcentuales. En ese caso, la brecha de género se achica, y el indicador Eika se acerca a 2. Finalmente, en el tercer período, la brecha se anula y únicamente en ese caso, el indicador reporta que el emparejamiento es perfecto, a pesar de que el patrón se mantuvo a lo largo de los tres períodos, es decir, a pesar de que los individuos pertenecientes al mismo ranking de cada distribución se emparejaron en los tres casos.

Tabla A.3.1. Crítica Indicador Eika

M1. Mat obs (Perfecta)				M1. Mat obs (Perfecta)				M1. Mat obs (Perf	
H\M	Alta	Baja	Total	H\M	Alta	Baja	Total	H\M	Alta
Alta	0.50	0.20	0.70	Alta	0.50	0.10	0.60	Alta	0.50
Baja	0.00	0.30	0.30	Baja	0.00	0.40	0.40	Baja	0.00
Total	0.50	0.50		Total	0.50	0.50		Total	0.50
M2. Matriz aleatoria				M2. Matriz aleatoria				M2. Matriz aleat	
H\M	Alta	Baja		H\M	Alta	Baja		H\M	Alta
Alta	0.35	0.35		Alta	0.30	0.30		Alta	0.25
Baja	0.15	0.15		Baja	0.20	0.20		Baja	0.25
M3. Matriz Eika				M3. Matriz Eika				M3. Matriz Eil	
H\M	Alta	Baja		H\M	Alta	Baja		H\M	Alta
Alta	1.43	0.57		Alta	1.67	0.33		Alta	2
Baja	0.00	2.00		Baja	0.00	2.00		Baja	0
Indicador Eika:		1.64		Indicador Eika:		1.81		Indicador Eika:	

A pesar de que este problema parece resolverse cuando se controlan las brechas de género (escenarios 3 y 4), al examinar la evolución de cada categoría por separado en relación con un escenario contrafactual aleatorio, los desafíos persisten. La Figura A.3.3 presenta los resultados de la evolución de las categorías de la diagonal principal de la tabla de contingencia a lo largo del tiempo.

Figura A.3.3. Tendencia del indicador Eika por categorías – Resultados de las simulaciones



Notas: la figura presenta las estimaciones del indicador de Eika por categorías para cuatro escenarios alternativos: el observado en la realidad y tres simulados. Es decir, para cada categoría educativa, se calculan las probabilidades (en relación a la coincidencia aleatoria) de observar una pareja en la que ambos miembros tienen ese nivel educativo. Los datos corresponden a los censos de Argentina.
Fuente: IPUMS-I (2020)

Los resultados evidencian que, incluso al considerar el control de las brechas de género, las categorías correspondientes al escenario observado y a las dos primeras simulaciones (escenarios 1, 2 y 3) muestran tendencias similares. Sin embargo, en la última simulación, donde se controla el tamaño de los grupos con UNIC y PRIC, se observa que las tendencias de dichas categorías se mantienen estables a lo largo del tiempo. Bratsberg et al. (2018) explican que estos resultados surgen como consecuencia de un efecto de composición de los grupos.

La composición de los grupos educativos ha cambiado necesariamente como resultado de la expansión educativa en todas las cohortes. Si bien los cambios mecánicos resultantes en las probabilidades de emparejamiento pueden controlarse normalizando las frecuencias de emparejamiento observadas con las que se habrían aplicado en el emparejamiento aleatorio, dicha normalización no se ocupa de los posibles cambios asociados en la composición de los grupos educativos.

Para entender la intuición, se presenta el siguiente ejemplo, representado en la tabla A.3.2. Considere una población de hombres y mujeres, divididos en dos niveles educativos (alto y bajo) en dos momentos diferentes del tiempo. En ambos casos, el patrón de emparejamiento es perfecto y no existen brechas de género en educación. Sin embargo, entre el primero y segundo período, se da una gran expansión educativa. Mientras que, en el primer caso, el 90% de los integrantes de cada género son poco educados, en la segunda distribución solamente el 50% es poco educado. Dado que el emparejamiento es perfecto y que no existen brechas de género, el indicador agregado de Eika es igual a 2. Sin embargo, cuando uno observa los valores de la matriz Eika por categoría, se observa que, dependiendo el tamaño de las distribuciones observadas, el contrafactual aleatorio hará que los valores finales de la matriz Eika crezcan de manera cuadrática a medida que disminuye el tamaño de los grupos.

En pocas palabras, si el logro educativo aumenta entre las cohortes, esto puede causar que el logro bajo identifique un grupo más pequeño y más homogéneo y el logro alto identifique un grupo más grande y heterogéneo en términos de alguna otra característica que puede ser el factor de emparejamiento real. Si este fuera el caso, observaríamos un emparejamiento selectivo educativo creciente para el grupo de baja educación y un apareamiento selectivo decreciente para el grupo de educación alta, incluso si los patrones de apareamiento subyacentes no cambiaran por completo.

En resumen, si bien el indicador propuesto por Eika et al. (2019) pretende controlar los cambios en las distribuciones mediante una normalización basada en un contrafactual aleatorio, no tiene en cuenta los posibles problemas que pueden surgir debido a los cambios en las brechas de género, ni considera las variaciones en la composición de cada grupo a medida que evolucionan las distribuciones.

Tabla A.3.2. Crítica Indicador Eika por categorías

M. Perfecta				M. Perfecta			
H\M	Alta	Baja	Total	H\M	Alta	Baja	Total
Alta	0.90	0.00	0.90	Alta	0.50	0.00	0.50
Baja	0.00	0.10	0.10	Baja	0.00	0.50	0.50
Total	0.90	0.10		Total	0.50	0.50	

Matriz aleatoria			Matriz aleatoria		
H\M	Alta	Baja	H\M	Alta	Baja
Alta	0.81	0.09	Alta	0.25	0.25
Baja	0.09	0.01	Baja	0.25	0.25

Matriz Eika			Matriz Eika		
H\M	Alta	Baja	H\M	Alta	Baja
Alta	1.10	0.00	Alta	2.00	0.00
Baja	0.00	10.00	Baja	0.00	2.00

Indicador Eika:	2	Indicador Eika:	2
-----------------	---	-----------------	---

Apéndice IV – Funcionamiento y crítica al indicador de Shen (2019)

Shen propone una métrica denominada "Normalización Perfecta-Aleatoria" (PRN, por sus siglas en inglés), que se define como una normalización que toma un valor de cero si el patrón de emparejamiento observado es aleatorio y un valor de uno si es perfecto. Luego, se evalúa dónde se encuentran los emparejamientos observados en relación con estos dos límites²⁰. El indicador se computa a partir de tablas de contingencia. Formalmente, sean E_o y E_a dos variables aleatorias para el nivel de educación de un hombre y una mujer, respectivamente, y sean e_o y e_a dos niveles de educación cualesquiera. Entonces, se denota $p_{\text{sim}}(e_o, e_a)$ donde $\text{sim} \in \{\text{Perfecta}, \text{Aleatoria}, \text{Observada}\}$ como la probabilidad de observar un emparejamiento con $E_o = e_o$ y $E_a = e_a$ en cada. La normalización perfecta aleatoria es:

$$s_{\text{Shen}}(e_o, e_a) = \frac{p_{\text{observada}}(E_o = e_o, E_a = e_a) - p_{\text{aleatoria}}(E_o = e_o, E_a = e_a)}{p_{\text{perfecta}}(E_o = e_o, E_a = e_a) - p_{\text{aleatoria}}(E_o = e_o, E_a = e_a)}$$

La simulación de emparejamiento aleatorio lo es en un sentido estadístico, es decir, asume que E_o y E_a son variables aleatorias independientes, entonces $p_{\text{aleatoria}}(e_o, e_a) = P(E_o = e_o) * P(E_a = e_a)$ ²¹. El problema del indicador de Shen radica en que puede fallar en situaciones donde la brecha de género es tan significativa en un nivel educativo que no es compensada por el residuo del siguiente.

El indicador de Shen tiene un buen funcionamiento, en el sentido de que los valores de la métrica se encuentran entre -1 y 1, para la mayoría de los países y los años considerados. Sin embargo, este parece mostrar fallas en algunos casos en los cuales las distribuciones educativas presentan grandes brechas de género. Un ejemplo de dicha falla se presenta al considerar el censo de Brasil correspondiente a la década de 1960. Brasil es un país cuyo desarrollo educativo ha sido muy pobre durante muchos años, principalmente a mediados del siglo XX. La distribución educativa de hombres y mujeres, al dividir en cuatro categorías, presenta la siguiente forma:

²⁰Si las parejas observadas están por debajo del contrafactual de emparejamiento aleatorio (ES negativo), entonces se trata al emparejamiento aleatorio como el límite superior (normalizado para que sea igual a 0) y se simula un escenario de emparejamiento perfectamente negativo, es decir, el hombre "top" se empareja con N-ésima mujer de la distribución educativa, el hombre 2 coincide con N-1, etc. Luego, se considera al contrafactual de emparejamiento perfectamente negativo como límite inferior, normalizado para ser -1, y la Normalización perfecta-aleatoria (para el ES negativo) pregunta dónde, entre estos dos límites, se encuentran los emparejamientos observados.

²¹ La expresión analítica para $p_{\text{perfecta}}(e, e')$ depende de cómo coincidan los hombres y las mujeres "sobrantes".

Tabla A4.1. Matriz observada distribución educativa parejas – Brasil, 1960.

Mujer/Hombre	PRII	PRIC	SECC	UNIC	Total
PRII	92.16	2.05	0.78	0.31	95.30
PRIC	1.04	1.06	0.57	0.51	3.19
SECC	0.35	0.29	0.41	0.33	1.38
UNIC	0.01	0.02	0.02	0.08	0.13
Total	93.56	3.43	1.78	1.24	100.00

En base a dicha tabla de contingencia observada, es posible calcular el contrafactual de emparejamiento perfecto de la manera en que lo propone Shen. A continuación, se presenta la tabla correspondiente a dicho escenario hipotético.

Tabla A.4.2. Matriz de emparejamiento perfecto – Brasil, 1960.

Mujer/Hombre	PRII	PRIC	SECC	UNIC	Total
PRII	93.56	1.74	0.00	0.00	95.30
PRIC	0.00	1.69	1.50	0.00	3.19
SECC	0.00	0.00	0.28	1.11	1.38
UNIC	0.00	0.00	0.00	0.13	0.13
Total	93.56	3.43	1.78	1.24	100.00

Se puede observar que en un caso como el presentado, en el cual las distribuciones educativas presentan grandes brechas en algunos niveles educativos, el valor observado en la diagonal principal es mayor al valor del contrafactual perfecto. Esto significa que el indicador de Shen tomará valores mayores a 1.

Al ver la tabla de contingencia del escenario contrafactual, se observa que la brecha en el nivel universitario entre hombres y mujeres es tal que los hombres universitarios se juntan con mujeres universitarias, y luego, el residual de hombres se junta con mujeres con educación secundaria completa. El problema es que es tan grande la diferencia entre universitarios (y tan parecida al porcentaje de individuos con secundario), que el residual de mujeres que queda para emparejarse con hombres es menor al porcentaje observado.

Apéndice V – Simulación Gihleb y Lang.

Siguiendo el enfoque presentado por Gihleb y Lang (2020), se examina el rendimiento de estas diferentes métricas, incluyendo la correlación lineal de Pearson, utilizando los datos de censos para parejas en Argentina. Para llevar a cabo el análisis, se simulan dos escenarios de emparejamiento perfecto. En el primero, se divide la distribución de años de educación en rangos de 0 a 18, teniendo en cuenta la disponibilidad de datos. En el segundo, se divide la distribución de educación en cuatro categorías educativas estandarizadas según el proyecto IPUMS. Los escenarios de emparejamiento perfecto se simulan de la misma manera que se describió en la subsección anterior, y a continuación se presentan los resultados correspondientes para la distribución de años educativos (los resultados relacionados con la estimación basada en niveles educativos se encuentran en la [Tabla A4](#)).

Tabla 2. Resultados de las diferentes métricas bajo emparejamiento perfecto.

Año	ρ	r	γ	τ_a	τ_b
1970	0.98	0.98	1.00	0.81	0.95
1980	0.99	0.99	1.00	0.87	0.97
1991	1.00	1.00	1.00	0.87	0.99
2001	0.99	0.99	1.00	0.82	0.96
2010	0.99	0.99	1.00	0.85	0.96

Nota: la tabla presenta los resultados que surgen al estimar diferentes métricas de correlación en un escenario en el que se simula emparejamiento perfecto, utilizando datos de Argentina para diferentes décadas de censo.

Fuente: IPUMS-I (2020)

Los resultados presentados en la Tabla 2 revelan que, al considerar los años de educación como variable de interés, tanto el coeficiente de correlación de Pearson como el coeficiente de correlación de Spearman muestran un desempeño similar y prácticamente perfecto. Sin embargo, cuando se examinan las cuatro categorías educativas, se observa que el coeficiente de correlación de rangos de Spearman exhibe un mejor rendimiento, seguido de cerca por la Tau B de Kendall y el coeficiente de correlación de Pearson. En contraste, la Tau A de Kendall presenta un desempeño deficiente en ambos casos. Es importante destacar que, debido al diseño del “experimento”, el indicador de Goodman y Kruskal es igual a 1 en todos los períodos.

Apéndice VI. Transformación de la variable educativa

La definición de la variable educativa transformada y de los cuantiles que surgen de esta se realiza teniendo en consideración dos aspectos fundamentales de las distribuciones de educación. En primer lugar, la educación se encuentra limitada en un rango entre 0 y 18 años, con puntos de acumulación correspondientes a la finalización de cada nivel educativo (primario, 6 años; secundario, 12; universitario, 17/18 años). Si se definieran cuantiles en función de esta variable, se correría el riesgo de incluir los mismos valores de años de educación en diferentes cuantiles. En segundo lugar, dada la expansión educativa en la región, existe una relación positiva entre el año de nacimiento y la educación promedio de los individuos. Si además se considera que las edades de los cónyuges son similares en la práctica, el hecho de no controlar por estos cambios podría generar distorsiones en las métricas de ES de cada censo.

Para definir cuantiles en base a una variable continua y permitir la comparación entre rankings de individuos de diferentes edades, se sigue la metodología propuesta por Halliday et al. (2021). Esta implica definir cuantiles en función de los residuos de una regresión de los años de educación en relación con la variable de edad. De manera formal, para cada censo disponible, se estima la siguiente regresión de Mínimos Cuadrados Ordinarios:

$$aedu_{id} = \alpha + \sum_{a \in A} \beta^a \text{año_nac}_{id}^a + \epsilon_{id} \text{ para } g[1,2],$$

donde $aedu_{id}$ representa los años de escolaridad del individuo i perteneciente al censo relevado en la década d . La variable independiente año_nac_{id}^a representa el año de nacimiento a del individuo i perteneciente al censo relevado en la década d . La regresión se corre para cada país y para cada género, g , por separado. Luego de correr cada regresión, se obtienen los residuos de las mismas y se definen cuantiles en base a estos.

Ventajas del ICEL

El Indicador Cuantílico Educativo Local, además de ser insensible a los cambios en las distribuciones marginales de educación, presenta una serie de ventajas que lo hacen apropiado para el estudio del ES. En primer lugar, permite controlar los cambios en las distribuciones de educación entre diferentes censos y entre cohortes de nacimiento dentro de un mismo censo, además de los cambios en las brechas de género y las variaciones en la composición de los grupos. En segundo lugar, la división permite estudiar el emparejamiento en diferentes partes de la distribución educativa. En tercer lugar, la interpretación del indicador es más intuitiva que, por ejemplo, los índices tradicionales de correlación. El cuarto punto se refiere al esquema de clasificación de las categorías educativas que se elija. Esta elección puede ser arbitraria y basarse en divisiones tradicionales, pero no implica que sea la única o la más adecuada. Además, diferentes esquemas de clasificación pueden dar lugar a diferentes tendencias (Gihleb & Lang, 2020).²² Al considerar cuantiles de la distribución, esta

²² Un contrapunto importante es que, al separar las categorías educativas de la manera tradicional, se reconoce que la educación también es una señal. Por ejemplo, tener 11 o 12 años de educación puede no parecer una diferencia significativa en términos de años, pero puede representar una distinción importante en términos de poseer o no un diploma de secundaria completa.

métrica omite dicha decisión. Por último, es posible generar extensiones directas a partir del mismo indicador, las cuales serán presentadas en las próximas secciones.

Apéndice VII. Metodologías de aleatorización para estudiar efectos distributivos

La literatura sobre emparejamiento selectivo ha estudiado de manera recurrente cómo cambiaría la desigualdad de ingresos de los hogares si las parejas se formaran aleatoriamente en lugar de seguir los patrones observados (Eika et al., 2019; Greenwood et al., 2014; Hryshko et al., 2017, para Estados Unidos; Leal, 2015; Pereira y Santos, 2017 para países latinoamericanos). En estos ejercicios, la aleatorización genera pseudohogares, es decir, hogares contrafactuales contruidos al reasignar parejas manteniendo fija la población y (según el caso) ciertas características observables.

Para abordar este punto, se han empleado dos enfoques alternativos de aleatorización conocidos como “aditivo” y “de imputación”, según la terminología propuesta por Harmenberg (2014). Ambos enfoques se orientan hacia la aleatorización de la formación de hogares, aunque divergen en la forma en que asignan los ingresos al nuevo hogar formado aleatoriamente –denominado “pseudohogar”. El enfoque aditivo calcula los ingresos de los pseudohogares sumando los ingresos individuales, mientras que el enfoque de imputación establece los ingresos del pseudohogar asumiendo que siguen la misma distribución que los ingresos reales de los hogares con características semejantes. Formalmente, sean x_i, x_j características individuales (edad, educación, etc.) de hombres y mujeres, y sean y_i, y_j los ingresos individuales de hombres y mujeres, respectivamente. El ingreso familiar de un hogar (i, j) , y_{ij} se define naturalmente como $y_{ij} = y_i + y_j$. Sea $dF(z)$ la distribución (verdadera) de una variable dada z . El enfoque aditivo, manteniendo constantes los atributos x_i, x_j en parejas, equivale a calcular la distribución:

$$\left(dF(y_i | x_i) * dF(y_j | x_j) \right) dF(x_i, x_j)$$

donde $*$ denota el operador de convolución. Es decir, se supone que a los agentes con características (i, j) se les asigna aleatoriamente una pareja, de modo que $y_i | x_i$ y $y_j | x_j$ son independientes. Bajo la aleatorización incondicional, los atributos que se mantienen constantes, x_i, x_j , están vacíos, por lo que esto se reduce a:

$$dF(y_i) * dF(y_j)$$

El enfoque de imputación, utilizando las características observables x_i, x_j , en su lugar calcula:

$$dF(y_{ij} | x_i, x_j) dF(x_i) dF(x_j)$$

Bajo este esquema, el ingreso del hogar se imputa a partir de las características observables del mismo. Estos dos enfoques de aleatorización son conceptualmente distintos y, en general, tiene poco sentido comparar los resultados que surgen de cada uno de ellos. El beneficio de la aleatorización por adición es que toma la población existente y la aleatoriza, con lo cual no hay pérdida de información. El gran inconveniente es que dicho método considera que la oferta de trabajo y los ingresos son exógenos a la formación del hogar. El beneficio de la

aleatorización de la imputación es que sí tiene en cuenta que la oferta laboral y los ingresos son endógenos a la formación del hogar. El inconveniente de dicha metodología en la práctica es que los atributos disponibles dan una imputación menos que perfecta.

