



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: I2, I24, J11, J12

15 de enero, 2025

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

Albina, Iván*

Maestría en Economía
Facultad de Ciencias Económicas
Universidad Nacional de La Plata

Directora: Lucila Berniell
Codirector: Leonardo Gasparini

* 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 Laguinde 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.

* Centro de Estudios Distributivos, Laborales y Sociales, Instituto de Investigaciones Económicas, Universidad Nacional de La Plata, Argentina. E-mail: ivan_albina@hotmail.com

Resumen

Este trabajo documenta el grado y la evolución del emparejamiento selectivo (ES) para cohortes nacidas entre 1915 y 1985 en 21 países de América Latina y el Caribe (LAC). Para ello, evalúa los indicadores de ES existentes, identifica los más robustos para medir su evolución controlando por variaciones en las distribuciones educativas y propone una nueva métrica que facilita el estudio del nivel de ES entre grupos de distintos niveles educativos, además de ser robusta a las variaciones en las distribuciones educativas. También, incluye dos extensiones de dicha métrica que consideran la selección potencial entre individuos que deciden formar una pareja y la dimensión geográfica en el análisis del ES agregado. Adicionalmente, presenta un ejercicio de simulaciones para cuantificar la contribución de distintos factores en el ES agregado en México, así como los cambios en la desigualdad bajo escenarios de emparejamiento aleatorio con diversas restricciones.

Códigos JEL: I2, I24, J11, J12

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

Contents

1. Introducción	3
2. Datos y estadísticas descriptivas.....	7
2.a. Fuentes de datos.....	7
2.b. Estadísticas descriptivas	8
3. Metodología	12
3. a. Indicadores de primera generación	12
3.b. Indicadores de segunda generación	13
3.c. Indicadores de tercera generación.....	14
3.d.1. Indicador Cuantílico Educativo.....	15
3.d.2. Extensiones del ICEL.....	17
4. Resultados	22
4.a. Emparejamiento Selectivo en la región	22
4.b. Evolución del ES en la región.....	23
4.c. Evolución del ES por quintiles.....	25
4.d. Indicador Cuantílico Educativo Global.....	26
4.e. Indicador Cuantílico Educativo Geográfico.....	29
5. Determinantes del emparejamiento	31
6. Emparejamiento selectivo y desigualdad	33
7. Comentarios finales	36
Referencias	38
Apéndice	41

1. Introducción

Durante la última década, el estudio del emparejamiento selectivo (ES) ha atraído renovada atención entre sociólogos y economistas. El grado de ES, definido como la tendencia de las personas a “formar pareja con los suyos”,¹² tiene potenciales implicancias sobre la desigualdad de ingresos y de otros recursos dentro de una sociedad. Asimismo, a través de la transmisión intergeneracional de atributos económicos, genéticos y socialmente heredables, este fenómeno puede perpetuar altos niveles de desigualdad y limitar la movilidad social en las generaciones futuras.

A pesar de la relevancia que puede tener el fenómeno para moldear la desigualdad, aún no se ha alcanzado un consenso sobre su evolución en los últimos años. Este resultado se explica, en parte, por el hecho de que la medición del mismo entraña dificultades de distinta naturaleza y aun no existe un acuerdo sobre la métrica más adecuada para evaluarlo (Bratsberg et al., 2023; Chiappori et al., 2021, 2025). El principal desafío es de índole metodológica, dado que los cambios en los patrones de formación de parejas pueden responder a múltiples factores, tales como variaciones en las proporciones de hombres y mujeres con distintos niveles de educación —variaciones en las distribuciones marginales de educación—, así como a cambios en factores asociados a la “tecnología de emparejamiento” (normas sociales, preferencias individuales, barreras en el mercado matrimonial que limitan conexiones entre personas, entre otros) (Becker, 1974). Para estudiar las implicancias que pueden tener los cambios en los patrones de emparejamiento sobre la desigualdad y la movilidad social resulta fundamental entender los cambios en el ES que surgen como consecuencia de cambios en la tecnología de emparejamiento, mientras que se intenta controlar por los cambios resultantes de la primera fuente de variación, los cambios en las distribuciones educativas. Este punto resulta crucial al realizar comparaciones a lo largo del tiempo y entre países, especialmente en América Latina y el Caribe, donde se han registrado importantes expansiones educativas desde mediados del siglo XX.³

Más allá de los desafíos asociados a la medición mencionados previamente, existen algunas dimensiones que no deberían ser pasadas por alto si se quisiera comprender el papel de los patrones de formación de pareja sobre la desigualdad. En primer lugar, los patrones de emparejamiento pueden diferir en los distintos niveles de la distribución educativa. En este sentido, si las parejas se forman principalmente por individuos pertenecientes a los extremos de la distribución educativa, las implicancias sobre la desigualdad serán distintas a las de un emparejamiento concentrado en los niveles intermedios. En segundo lugar, es relevante considerar que las personas que deciden formar pareja pueden diferir en características importantes de aquellas que no lo hacen. Incluir esta potencial selección es clave para medir la evolución del fenómeno,

¹ El emparejamiento selectivo positivo hace referencia a la tendencia de individuos con rasgos similares a formar pareja con mayor frecuencia de lo que se esperaría bajo un patrón de emparejamiento aleatorio. Típicamente se estudia este fenómeno en términos de la capacidad generadora de ingreso de los cónyuges. Por ejemplo, es común estudiar ES en función de la educación de las personas.

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 conformado 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.

³ Mientras el promedio de años de educación en 1960 ascendía a 3.57 años y en 3.27 años para hombres y mujeres de entre 25 y 45 años, esta cifra pasó a ser de 10.5 y 10.7, respectivamente, en 2010.

especialmente en la región, donde se ha registrado una caída significativa en la proporción de personas en pareja —fenómeno conocido como *unmating*— en los últimos 50 años.⁴ Finalmente, los patrones de emparejamiento pueden variar según la región geográfica, añadiendo un elemento adicional de complejidad a la medición y comprensión del fenómeno. En América Latina, una región caracterizada por grandes desigualdades espaciales, baja movilidad interna y amplias extensiones geográficas (De La Mata et al., 2022), esta dimensión adquiere especial relevancia.

Este trabajo aborda los desafíos mencionados mediante cuatro contribuciones al estudio del emparejamiento selectivo.

En primer lugar, se presentan y se evalúan las principales métricas utilizadas en la literatura de ES y, en base a las fallas identificadas en algunas ellas, se propone un nuevo indicador, denominado “Indicador Cuantílico Educativo Local” (ICEL). El ICEL mide la probabilidad de que dos individuos pertenecientes al mismo cuantil educativo formen pareja, en comparación con la probabilidad esperada en un escenario de emparejamiento aleatorio en términos educativos. Este indicador es particularmente adecuado para analizar la evolución del fenómeno, ya que permite explorar patrones entre diferentes cuantiles, es robusto frente a cambios en las distribuciones educativas y en las brechas de género, y proporciona una interpretación clara de los niveles de ES, por país y a lo largo del tiempo.

En segundo lugar, se presentan dos posibles extensiones del ICEL que amplían su utilidad. La primera está representada por el “Indicador Cuantílico Educativo Global” (ICEG), que incorpora en el resultado la posible selección en características que podrían existir entre quienes deciden formar pareja, además de considerar la selección entre quienes efectivamente forman una. La segunda extensión, el “Indicador Cuantílico Educativo Geográfico” (ICE Geo), considera que no solo interesan los patrones de emparejamiento de una región, sino también el nivel educativo promedio de la misma para entender las implicancias de este sobre la desigualdad a nivel nacional.

En tercer lugar, en base a los indicadores tradicionales identificados como los más robustos, y a las nuevas propuestas, se estudia la evolución del ES en 21 países de América Latina y el Caribe. Para ello, se utiliza información proveniente de censos de población y vivienda estandarizados por el proyecto *IPUMS-International* (IPUMS, 2020)⁵, recogidos durante seis períodos (desde la década de 1960 hasta la de 2010, inclusive) correspondientes a cohortes nacidas entre 1915 y 1985.

La evidencia presentada motiva y guía una cuarta contribución: cuantificar el poder explicativo que tienen diferentes factores —geográficos, etarios y maritales— para explicar el nivel de emparejamiento selectivo agregado en un país, así como estimar los cambios en la desigualdad al asumir emparejamiento aleatorio incluyendo diferentes restricciones.

El trabajo ofrece tres grupos de resultados, que se comentan a continuación.

Primero, el análisis de las diferentes métricas muestra que algunas presentan limitaciones para estudiar los cambios en el fenómeno a lo largo del tiempo. En particular, algunas de estas fallan al intentar controlar los cambios provenientes de las variaciones

⁴ 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.

⁵ El uso de este tipo de información conlleva algunos beneficios adicionales que serán oportunamente descritos.

en las distribuciones marginales de educación. Los indicadores de correlación de rangos, así como el ICEL, se identifican como los indicadores más robustos para medir la evolución del fenómeno. En base a estos, se presenta un análisis del nivel y las tendencias de ES en 21 países de la región. Los resultados para las cohortes más recientes presentan evidencia de ES positivo en todos los países, con una gran heterogeneidad entre ellos. En promedio, el ICEL muestra que las probabilidades de que dos individuos del mismo cuantil educativo se emparejen son 2.3 veces mayores a las que surgirían bajo un patrón de emparejamiento, mientras que el coeficiente de correlación de rangos asciende a 0.63 para la región. Entre los países con menor ES figuran Puerto Rico, Uruguay y República Dominicana; mientras que, en el otro extremo, aparecen Haití, Guatemala y Chile. A su vez, se observa que el patrón de ES para los quintiles extremos es mayor al de los niveles medios de la distribución, con valores promedio de 2.5 y 2.97 para los individuos del primer y quinto quintil, respectivamente. En términos de la evolución del fenómeno, si bien se observan tendencias relativamente estables a lo largo del tiempo, al comparar los primeros registros de cada país con los últimos disponibles, se aprecia una caída en los valores del ES, en entre 13 y 15 de los 21 países, dependiendo del indicador considerado. Los resultados son consistentes con los obtenidos en otros estudios que han documentado el fenómeno (CEDLAS y Banco Mundial, 2024)⁶, y en conjunto, brindan un panorama que muestra que entre 1960 y 2020 hubo una caída generalizada en los patrones de emparejamiento en la mayoría de los países de la región.

En segundo lugar, se estudian los resultados basados en las extensiones del indicador. Al considerar a todo el universo de potenciales parejas —todos los individuos de entre 25 y 45 años— se observa que las mujeres que deciden emparejarse tienen un nivel educativo promedio menor que aquellas que no lo hacen, mientras que el patrón por parte de los hombres en pareja no es claro. A su vez, se observa que los hombres en pareja, en promedio, se ubican en un decil mayor —respecto a la distribución de su propio género— a las mujeres. Al comparar el ICEL con el ICEG, no se observan grandes diferencias entre los indicadores, pero se presenta evidencia de que, en la mayoría de los países, el indicador que incluye la selección de quienes deciden emparejarse tiende a presentar valores mayores a aquel que considera únicamente el universo de individuos en pareja. Por otro lado, las extensiones que consideran la dimensión espacial muestran que el nivel de emparejamiento varía significativamente entre regiones dentro de los países, y que las desigualdades geográficas en términos educativos llevan a que las probabilidades de que dos personas pertenecientes a los extremos de la distribución educativa se junten varíen fuertemente según la región. Entre las ciudades con mayor selección en el quintil más educado de América Latina y el Caribe se encuentran: Antiguo Cuscatlán, Nuevo Cuscatlán y Santa Tecla (en La Libertad, El Salvador); Vitacura, Providencia, Las Condes y Ñuñoa (en Santiago, Chile), San Pedro Garza García (en Nuevo León, México), Asunción (Paraguay), San Andrés Huayápam (Oaxaca, México) y Envigado (Antioquía, Colombia).

Por último, al estudiar el peso de diferentes dimensiones —estado de emparejamiento, edad y geografía— sobre el emparejamiento en uno de los países más grandes de la región en dos momentos diferentes, México en el 2000 y en el 2020, los resultados indican que la geografía puede llegar a explicar hasta un sexto del emparejamiento observado y, que, en su conjunto, las diferentes dimensiones pueden llegar a explicar hasta un tercio del mismo. En esta línea, al estudiar la contribución del emparejamiento sobre la desigualdad de este país, los resultados sugieren que, si bien se evidencia una fuerte disminución en

⁶ Estos últimos, sin embargo, cubren períodos más cortos y se basan en encuestas de hogares.

el nivel de desigualdad al considerar el emparejamiento aleatorio (caídas de entre el 5.8% y el 6.4% del Gini), la incorporación de las diferentes restricciones en los contrafacticos mitiga parte de dicha reducción (que pasa a ser de, aproximadamente, el 3%).

Este trabajo se relaciona con un cuerpo de literatura empírica que ha medido el ES en un período que abarca desde la Segunda Guerra Mundial hasta la década de 2010 en países desarrollados, principalmente en Estados Unidos. En esta literatura, se han documentado diversas tendencias de emparejamiento selectivo según la métrica utilizada, lo que ha generado debates sobre el mejor indicador para medir la evolución del fenómeno. La literatura inició cuantificando el ES sin tener en cuenta el papel que pueden tener los cambios en las distribuciones marginales educativas sobre las métricas de ES (Breen & Salazar, 2011; Fernandez et al., 2005; Ganguli et al., 2014; Greenwood et al., 2014; Kremer, 1997) y continuó con otras propuestas que intentaron controlarlos, pero sin completo éxito (Eika et al., 2019; Schwartz & Mare, 2005; Shen, 2019, entre otros). Más recientemente, surgió una tercera generación de métricas que ostenta un mejor desempeño en este sentido (Bratsberg et al., 2023; Chiappori et al., 2021; Gihleb & Lang, 2020; Handy, 2021). Este documento expande los trabajos previos en varias direcciones. Primero, evalúa las métricas existentes de manera sistemática, identificando su capacidad para controlar por los cambios en las distribuciones educativas. En segundo lugar, presenta evidencia para un grupo de 21 países de la región por un período extenso, revelando diferentes tendencias y niveles de emparejamiento. En tercer lugar, la propuesta de una nueva métrica permite estudiar los patrones de emparejamiento entre los diferentes cuantiles educativos.

El trabajo también contribuye a una literatura que intenta dar cuenta de los cambios en la desigualdad económica observados en países de América Latina a partir de las variaciones en los patrones de emparejamiento selectivo observados. Esta dimensión ha recibido notable atención en literatura previa (Ciscato & Weber, 2020; Eika et al., 2019; Greenwood et al., 2014; Harmenberg, 2014, para Estados Unidos, y Funes Leal, 2015; González & Dip, 2022 para Argentina, entre otros). Este trabajo expande la evidencia para el caso mexicano y pone especial énfasis en las restricciones consideradas en cada escenario de emparejamiento aleatorio.

El resto de este documento está organizado de la siguiente forma. En la sección 2, se presentan las fuentes de datos y algunas estadísticas descriptivas. La sección 3 describe los desafíos de la medición y la metodología utilizada. En la sección 4, se presentan los resultados del ES en la región. La sección 5 estudia el peso relativo de los diferentes factores sobre el emparejamiento, mientras que la 6 evalúa el impacto que tiene el ES en la desigualdad, una vez consideradas restricciones etarias, de estado conyugal y de residencia. Por último, la sección 7 concluye y propone futuras líneas de investigación.

2. Datos y estadísticas descriptivas

2.a. Fuentes de datos

Para estudiar el nivel y la evolución del fenómeno en la región, se utilizan muestras de censos de población armonizadas por el proyecto IPUMS *International* (IPUMS, 2020) correspondientes a 21 países de América Latina y el Caribe. 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.⁷⁸ Los datos corresponden a censos que parten de la década de 1960 y llegan hasta la de 2010 (en la [Tabla A1](#) se presentan los censos utilizados de cada uno de los países).

Las variables capturadas en IPUMS-I incluyen información demográfica individual (género, fecha de nacimiento, estado civil, etc.) y datos socioeconómicos (nivel educativo, estado ocupacional e ingresos, los últimos dos solo para algunos países). El nivel educativo se observa a través de dos variables armonizadas, lo que permite la comparación internacional. La primera registra el máximo nivel educativo alcanzado por los individuos, mientras que la segunda reporta el número de años de educación formal recibida, comenzando desde la primaria.⁹ Los datos censales, a su vez, contienen identificadores familiares únicos que permiten emparejar a los cónyuges de un mismo hogar. El [Apéndice I](#) se refiere al método utilizado para vincular parejas en este trabajo.

Para estudiar el emparejamiento selectivo a nivel regional, se emplean dos variables geográficas disponibles en la base de datos: *geolev1* y *geolev2*. Estas identifican unidades administrativas subnacionales de primer y segundo nivel, respectivamente, que mantienen la estabilidad geográfica a lo largo del tiempo. 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.¹⁰

Los censos representan la fuente de información estadística más importante y extensa disponible para los países, y presentan tres características que los hacen especialmente adecuados para el estudio del emparejamiento selectivo educativo: su rango temporal – que abarca hasta 50 años—; la cobertura nacional y la desagregación espacial; y, por último, el número de observaciones proporcionadas. Estas últimas dos características permiten estimar métricas de ES en pequeñas regiones geográficas, las cuales pueden representar una aproximación precisa a los diferentes mercados matrimoniales dentro de cada país.

⁷ 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.

⁸ 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.

⁹ El máximo de valor que pueden tomar los años de educación alcanzados puede variar en algunos países. [Tabla A2](#) en el apéndice.

¹⁰ Es importante tener en cuenta que la superficie abarcada por cada unidad administrativa armonizada puede variar sustancialmente entre naciones. Por ejemplo, en México, la mediana del área de las unidades geográficas de segundo nivel es de 234 km², mientras que en Bolivia esta cifra asciende a 5,500 km². Estas diferencias en las superficies pueden ejercer una influencia significativa en los patrones de emparejamiento selectivo cuando se controla por el factor geográfico, por lo que resulta imperativo considerarlas al hacer el análisis.

2.b. Estadísticas descriptivas

El análisis principal se centra en parejas que cohabitan en América Latina y el Caribe, en las que al menos uno de los cónyuges tiene entre 25 y 45 años. En la Tabla 1, se presentan algunas estadísticas clave para el total hombres y mujeres dentro de este rango etario, dividiendo la muestra entre aquellos que se encuentran pareja y quienes permanecen solteros o separados¹¹ (en la [Tabla A3](#), se presentan estadísticas para todas las décadas).

La Tabla 1 muestra tres puntos clave. En primer lugar, se destaca la notoria expansión educativa en América Latina y el Caribe durante el siglo XX, con un aumento sustancial en el promedio de años de educación, explicado por una disminución en la proporción de personas con educación primaria incompleta, junto con un incremento en niveles educativos superiores restantes.¹² En segundo lugar, se observa que, si bien la expansión educativa fue generalizada en la región, las mujeres experimentaron un aumento mayor en su logro educativo en comparación con los hombres, reduciendo la brecha e incluso revirtiéndola en las últimas décadas.¹³ En tercer lugar, se evidencia una disminución en el porcentaje de personas en pareja a través del tiempo —fenómeno denominado *unmating*—, y se observa que las mujeres que tienden a formar pareja, en promedio, tienen menor nivel educativo en comparación a aquellas que no lo hacen. Esto resulta en una reversión de la brecha educativa más pronunciada entre las mujeres solteras.

¹¹ Si bien el análisis de parejas puede incluir a individuos mayores o menores a dicho rango etario, los datos muestran que el 85% de las mujeres y el 81.2% de los hombres tienen esa edad.

¹² Este fenómeno ha sido descrito en la literatura, incluyendo trabajos de Berniell et al. (2016, 2021).

¹³ A pesar de las variaciones en los países analizados y las tendencias cambiantes, el análisis por país confirma las mismas conclusiones. La [Figura A1](#) 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.

Tabla 1. Estadísticas descriptivas, según estado de pareja.

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

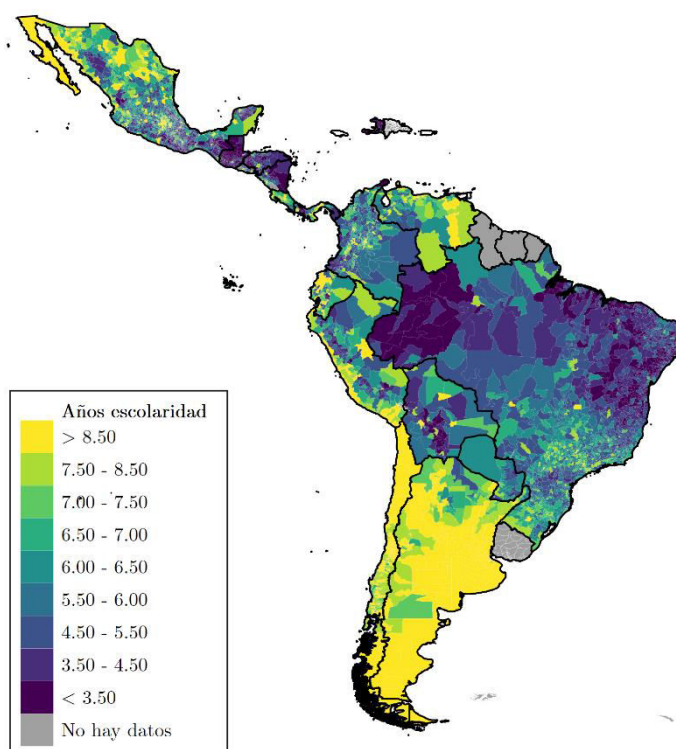
Nota: La tabla presenta el promedio regional de medias muestrales de edad y educación, así como también el total de observaciones para cada década censo. Los niveles educativos presentados en la tabla son Primaria Incompleta (PRII), Primaria Completa (PRIC), Secundaria Completa (SEC) y Universitario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

Fuente: IPUMS-I (2020)

La Figura 1 complementa las estadísticas proporcionadas en la Tabla 1 al ofrecer evidencia sobre la distribución espacial de la educación. Esta presenta el promedio de años de educación por unidad administrativa de segundo orden para los países con información disponible en la década del 2000.¹⁴ Los datos revelan una gran heterogeneidad educativa entre países —siendo Argentina, Chile y Costa Rica los más educados—, así como también dentro de cada ellos. Por ejemplo, en países como México y El Salvador, las unidades administrativas más instruidas exhiben promedios de años de educación formal que son hasta diez u once veces superiores a los de las menos educadas.

¹⁴ 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.

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

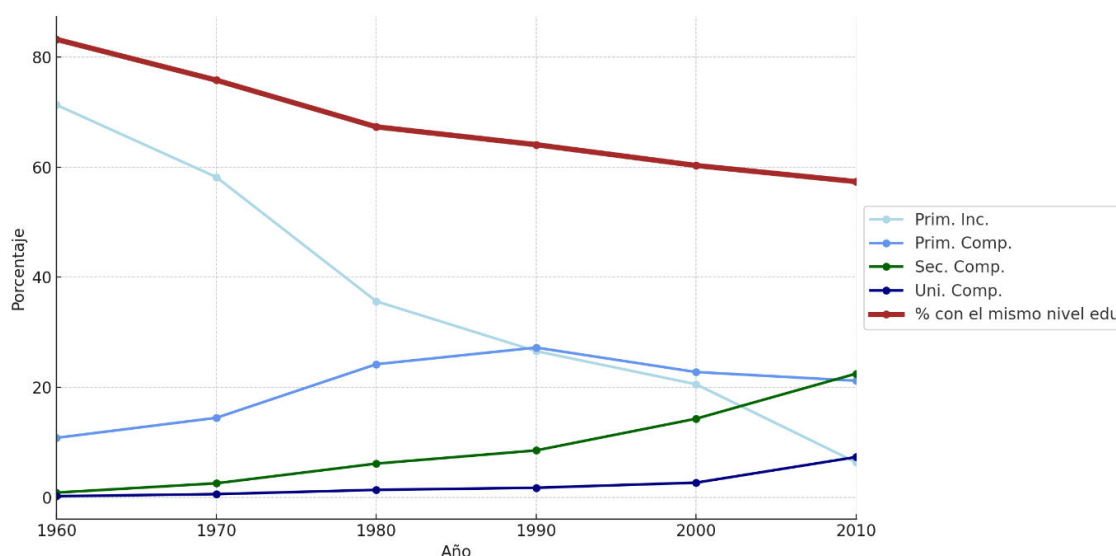


Nota: esta figura presenta los años de educación promedio para cada unidad administrativa de segundo orden para aquellos países que disponen de censos poblacionales en la década de los 2000.

Fuente: IPUMS-I (2020).

La evolución reportada en el nivel educativo promedio tiene su correlato con la transformación de la composición de las parejas a lo largo del tiempo. En la Figura 2 se presenta la proporción total de parejas homogámicas en términos educativos, es decir, parejas en las que ambos integrantes tienen el mismo nivel educativo, así como también la proporción desagregada por nivel, para el promedio de países de América Latina y el Caribe (en la [Figura A2](#), se presenta la evolución por país). En esta, se observa que los cambios en las distribuciones marginales educativas de la región han sido acompañados por variaciones en los porcentajes de parejas con el mismo nivel educativo, tanto dentro de cada nivel, así como en el total de parejas. Mientras que la proporción de parejas en las que ambos cónyuges tienen primaria incompleta disminuyó significativamente entre 1960 y 2010, los porcentajes de parejas homogámicas en los tres niveles superiores restantes registraron un aumento. En suma, dado que la caída en el porcentaje de parejas con primaria incompleta no llegó a ser compensada por el aumento del resto de los grupos, estos cambios dieron lugar a una caída en el porcentaje del total de parejas con el mismo nivel educativo con el paso de los años.

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 mecánica entre los cambios observados en las distribuciones educativas y los patrones de homogamia. La fuerte caída en la proporción de individuos con nivel primario incompleto parece haber tenido implicancias en la proporción de parejas con este nivel, y como consecuencia, en las tendencias agregadas de homogamia educativa. En base a las relaciones mencionadas, surge la inquietud acerca de cómo habrían cambiado los patrones de emparejamiento si se hubieran mantenido constantes las distribuciones marginales de educación. Es decir, se busca descomponer los cambios en los patrones de emparejamiento en dos: aquellos que responden a cambios mecánicos –los cuales surgen de las variaciones en las distribuciones educativas–, y aquellos correspondientes a la tecnología o a los precios de emparejamiento, siendo estos últimos los que tienen una mayor relevancia a la hora de entender su relación con la desigualdad de ingresos.

Además, las estadísticas muestran una tendencia decreciente en el porcentaje de parejas conformadas en la región, así como una gran heterogeneidad en términos educativos entre regiones. Por ese motivo, también se considera relevante la incorporación de dos nuevas dimensiones –quiénes deciden formar pareja y a qué región pertenecen– en el estudio del nivel y evolución del emparejamiento selectivo.

3. Metodología

La revisión bibliográfica de la introducción, así como el ejemplo presentado en la sección anterior, dejaron en evidencia la existencia de diversas métricas para documentar el nivel y la evolución del ES. Antes de presentar los indicadores que serán utilizados en este trabajo, se realizará un breve repaso de las principales métricas que han sido utilizadas en el estudio del ES, las cuales serán divididas en tres grupos —denominados “generaciones”— de acuerdo a la consideración que les han dado a los cambios en las distribuciones educativas y el éxito que han tenido al hacerlo. De esta manera, quedará en evidencia que los indicadores elegidos para el resto del análisis sortean las dificultades encontradas en la literatura.

3. a. Indicadores de primera generación

Dentro de la primera generación de indicadores, se encuentran aquellos que estudian los patrones de formación de parejas sin distinguir entre los cambios que ocurren como resultado de variaciones en las distribuciones educativas y aquellos que son resultado de cambios en la tecnología de emparejamiento.

Uno de los métodos más simples en esta categoría, presentado en la sección anterior, consiste en calcular el porcentaje de parejas que comparten el mismo nivel educativo o un nivel de diferencia, o alternativas similares —por ejemplo, aquellas que se presentan a través de métricas normalizadas.¹⁵ Más allá de la forma que tomen, todas estas métricas operan de manera similar, y tal como quedó evidenciado en la sección anterior, son sensibles a los cambios en las distribuciones educativas, así como también a la categorización que se utilice para dividir a la distribución educativa (Gihleb & Lang, 2020).

Un segundo indicador dentro de este grupo es el coeficiente de las regresiones de Mínimos Cuadrados Ordinarios (β_{MCO}), el cual ha sido ampliamente utilizado en la literatura (Greenwood et al., 2014, para Estados Unidos; Gabrielli y Serio, 2017 y González y Dip, 2022, para Argentina, entre otros). Este método consiste en estimar una regresión en la que la variable dependiente es la educación de la esposa y la independiente, la del esposo. Sin embargo, como explican Eika et al (2019), este indicador no es apropiado debido a que la métrica es sensible no solo a los cambios en las distribuciones educativas, sino también a la forma en que se establece el orden de las variables dependientes e independientes en la regresión.¹⁶ En el [Apéndice II](#) de este trabajo se proporciona un ejemplo basado en datos de Argentina que ilustra la invalidez de este indicador para medir la evolución del ES.

¹⁵ 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.

¹⁶ 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).

Por último, el coeficiente de correlación lineal de Pearson, ampliamente utilizado en la literatura (CEDLAS & World Bank, 2022; Fernandez et al., 2005; Kremer, 1997) también se puede ubicar dentro de los indicadores del primera generación, ya que mide la relación lineal, y no de rangos, entre dos variables cuantitativas. En otras palabras, la definición teórica del ES asume que este será perfecto cuando el hombre más educado se empareje con la mujer más educada, el segundo más educado lo haga con la segunda más educada, y así sucesivamente en orden descendente para cada distribución educativa. Esta definición no necesariamente implica una relación lineal entre los años de educación, sino que la mejor métrica debería ser una que estudiara la relación de rangos. Si bien más adelante se mostrará que el coeficiente de Pearson representa una métrica robusta para medir el fenómeno, este podría verse afectado tanto por cambios en las distribuciones educativas, así como por variaciones en los patrones de emparejamiento.

3.b. Indicadores de segunda generación

En el ámbito de los indicadores de segunda generación, se presentan algunos que han identificado el problema que pueden generar los cambios en las distribuciones marginales en educación a lo largo del tiempo pero que, aun así, no han logrado controlarlos por completo. En un esfuerzo por abordar este desafío, se han basado en la comparación de las parejas observadas en los datos con aquellas que surgirían en un mundo hipotético en el que hombres y mujeres se emparejan aleatoriamente en términos educativos. De acuerdo con esta lógica, a medida que los individuos de cada género adquieren más educación, el contrafactual de emparejamiento aleatorio cambiará en consecuencia, y dicha normalización medirá los cambios en el emparejamiento neto que habría ocurrido por casualidad aleatoria. Los métodos más destacados dentro de este grupo son los ratios de probabilidades (*odds ratios*); los modelos log-lineales (Mare, 2001; Schwartz & Mare, 2005; Bouchet Valat, 2014), ampliamente utilizados en la sociología; así como también el indicador propuesto en el trabajo de Eika et al. (2019). Si bien estos indicadores pretenden controlar los cambios en las distribuciones mediante la normalización basada en un contrafactual aleatorio, no tienen en cuenta los posibles problemas que pueden surgir debido a los cambios en las brechas de género, ni consideran las variaciones en la composición de cada grupo a medida que evolucionan las distribuciones (Bratsberg et al., 2023; Shen, 2019). En el [Apéndice III](#), se presenta detalladamente cada uno de los indicadores y se realiza un ejercicio de microsimulaciones, inspirado en el trabajo de Shen (2019), para mostrar algunos de los inconvenientes que pueden surgir al utilizar estos indicadores para medir la evolución del fenómeno.

Basándose en algunas de las críticas mencionadas previamente, el trabajo de Shen (2019) propone una métrica denominada “Normalización Perfecta-Aleatoria”, la cual estudia el nivel de emparejamiento selectivo observado en relación a un escenario de emparejamiento aleatorio, pero normalizando la métrica por un valor que surge de la diferencias entre el nivel de emparejamiento que se observaría si este fuera perfecto y el que se observaría si fuera aleatorio. Luego, se evalúa dónde se encuentran los emparejamientos observados en relación con estos dos límites. De esta manera, pretende estudiar la evolución del ES controlando los cambios en las brechas educativas y los cambios de composición de los grupos, dos de las fallas identificadas en los indicadores previos. Si bien en la mayoría de los casos en los que las distribuciones educativas de hombres y mujeres no presentan grandes diferencias el indicador funciona adecuadamente, este puede fallar en casos en los que las distribuciones educativas son

altamente desiguales. En el [Apéndice IV](#), se presenta la métrica en detalle y un ejemplo de la falla mencionada.

En resumen, la segunda generación de indicadores está conformada por métricas que reconocen la influencia de cambios en las distribuciones educativas en la medición del emparejamiento selectivo, intentan controlarlos, pero no logran hacerlo con éxito. Como respuesta a estas limitaciones, surge una tercera generación de indicadores que busca abordar las deficiencias identificadas.

3.c. Indicadores de tercera generación

Los indicadores de tercera generación se definen como aquellos que reconocen la necesidad de considerar los posibles inconvenientes que pueden surgir debido a los cambios en las distribuciones de educación al estudiar la evolución del ES. Estos identifican que dichos cambios no se limitan únicamente a las variaciones en la media de años de educación, sino que también pueden abarcar cambios en las brechas de género y en la composición de cada grupo educativo.

Tal como se comentó previamente, el ES es perfecto si el hombre con mayor educación se empareja con la mujer más educada, luego el segundo más alto en cada grupo se empareja, y así sucesivamente. En esa línea, Gihleb & Lang (2020) sugieren que lo más apropiado sería utilizar una medida de correlación diseñada específicamente para datos ordenados, que no dependa de las propiedades de intervalo de estos datos, es decir, una métrica que evalúe la correlación de rangos en lugar de la correlación lineal entre los años de educación. No obstante, destacan que las métricas de correlación de datos ordenados pueden presentar dificultades en presencia de “*ties*”¹⁷ (empates), los cuales son frecuentes en los datos de educación. A través de diferentes simulaciones, los autores evalúan las diferentes métricas que estudian la correlación de rangos y muestran que, en su mayoría, estas son efectivas para estudiar el emparejamiento selectivo. En particular, encuentran un buen funcionamiento en el Coeficiente de Correlación de rangos de Spearman, la Tau B de Kendall, el indicador de Goodman y Kruskal, así como también el coeficiente de correlación lineal de Pearson, a pesar de que este último no mida estrictamente correlación de *rankings*.¹⁸ La única excepción está dada por la Tau A de Kendall. En el [Apéndice V](#) de este trabajo, se practica el mismo ejercicio de simulaciones en base a los datos de Argentina y se obtienen conclusiones similares.

Un aspecto vinculado con los indicadores de correlación tiene que ver con que estos miden la relación entre los *rankings* educativos de los miembros de las parejas y que dichos *rankings* suelen estar basados en la distribución de todas las cohortes de cada género, para cada censo. Por lo tanto, si se producen cambios en el promedio de años de educación a lo largo del tiempo entre cohortes de un mismo censo, es probable que los individuos más jóvenes pertenezcan a *rankings* que, en promedio, sean más altos en términos educativos. Además, es común que las edades de los cónyuges sean similares en la práctica, lo que significa que los indicadores que no consideran estos cambios tienden a

¹⁷ Los *ties* son las observaciones dentro en una distribución que no son concordantes ni discordantes. En el entorno del emparejamiento selectivo, se dice que un par de parejas es concordante si tanto la esposa como el esposo de una pareja tienen una educación superior a la de la esposa y el esposo de la otra pareja. El par de parejas es discordante, en cambio, si una pareja tiene una esposa con menor educación y un esposo con mayor educación en comparación con la otra pareja.

¹⁸ 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.

sobreestimar el nivel de emparejamiento en cada año. En su lugar, una alternativa es utilizar rangos dentro de las distribuciones educativas específicas para cada género y cohorte de nacimiento, de modo que se tenga en cuenta que un nivel de logro educativo particular puede representar una posición relativa en la distribución educativa diferente dependiendo de la cohorte de cada individuo (Handy, 2021). A pesar de que las diferencias resultantes no cambian las conclusiones, vale tener en cuenta este punto para el indicador que se presentará a continuación.

Un último punto vinculado con estos indicadores radica en que, si bien son robustos para medir la evolución del emparejamiento selectivo, resulta también relevante analizar las tendencias de emparejamiento en cada nivel educativo, más allá del nivel agregado, debido a sus potenciales implicaciones en la desigualdad y la movilidad. En este sentido, el trabajo de Bratsberg et al. (2023) propone una métrica inspirada en el indicador de Eika (2019), pero basada en cuantiles de diversas variables socioeconómicas continuas en lugar de grupos educativos. En particular, los autores miden el grado de ES a través de un indicador que se define de la siguiente forma:

$$s(r_h, r_m) = \frac{P(R_h = r_h, R_m = r_m)}{P(R_h = r_h)P(R_m = r_m)}$$

donde R_h (R_m) denota el cuantil en la distribución de recursos del hombre (mujer). De este modo, para cada cuantil de la distribución de recursos, el indicador se define como la probabilidad de que un hombre del cuantil r_h se empareje con una mujer del cuantil r_m , en relación a la probabilidad de que esto sucediera si el patrón de emparejamiento fuera aleatorio en términos de recursos. Dado que los grupos definidos surgen de cuantiles, esta métrica, por definición, mantiene fijos los porcentajes de cada grupo en las distribuciones de cada género a lo largo del tiempo, evitando así uno de los principales problemas que surgen en el cálculo del indicador de Eika —variaciones en las brechas de género y en la composición de cada grupo educativo—, además de los cambios en la educación promedio. Este indicador tiene la ventaja de estudiar el emparejamiento para diferentes partes de la distribución, tal como lo hace el de Eika, pero sorteando las fallas de dicha métrica.

3.d.1. Indicador Cuantílico Educativo

En esta subsección, se propone una adaptación al indicador de Bratsberg et al. (2023) que se denominará “Indicador Cuantílico Educativo Local” (ICEL). La principal diferencia en relación a la métrica original radica en que, en lugar de considerar cuantiles definidos en base a variables vinculadas con el ingreso o la riqueza, considera una variable educativa transformada a los fines de la medición. Esta variable está representada por una métrica educativa continua y neta del efecto tendencia que surge de la expansión educativa en cada país. Nuevamente, en base a las tablas de contingencia, es posible medir el ES entre los cuantiles educativos ce_m y ce_h como la probabilidad observada de que un hombre del cuantil educativo ce_h esté en pareja con una mujer del cuantil educativo ce_m , en relación con la probabilidad bajo emparejamiento aleatorio con respecto a la educación.

Formalmente, el ICEL se define de la siguiente manera:

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

donde $CE_h(CE_m)$ denota el cuantil educativo correspondiente a cada cohorte de nacimiento del hombre (mujer). Así, la magnitud de $ICEL(ce_m, ce_h)$ se puede interpretar como la probabilidad de un emparejamiento entre individuos de dos cuantiles educativos (similares o diferentes entre sí) en comparación con la que surgiría bajo un patrón de emparejamiento aleatorio. En este caso, dado que se utilizan cuantiles, el denominador es consistente en todas las celdas y depende del cuantil considerado al dividir las distribuciones educativas. Por ejemplo, si se utilizan deciles, la matriz de coincidencia aleatoria consistirá en celdas con un valor de 0.01 ($0.1 \cdot 0.1$)¹⁹, por definición. Si se optara por quintiles, la matriz aleatoria contendrá celdas con un valor de 0.04. Para obtener una medida general de emparejamiento selectivo, se calcula el promedio ponderado de los parámetros de emparejamiento a lo largo de la diagonal principal de la matriz de transición educativa de hombres y mujeres. El emparejamiento selectivo positivo (negativo) significa que los hombres y mujeres con cuantiles educativos ce_h y ce_m forman pareja con mayor (menor) frecuencia de lo que se esperaría bajo un patrón de emparejamiento que es aleatorio en términos de educación, es decir, el $ICEL(ce_m, ce_h)$ es mayor (menor) que 1.

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:

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

¹⁹ 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 $0.1 \cdot 0.1$.

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).²⁰ Al considerar cuantiles de la distribución, esta métrica omite dicha decisión. Por último, es posible generar extensiones directas a partir del mismo indicador, las cuales serán presentadas en las próximas secciones.

3.d.2. Extensiones del ICEL

Indicador Global

La mayoría de la evidencia empírica se enfoca en el estudio de los patrones de emparejamiento entre individuos que declaran estar en pareja o cohabitar. Es decir, se calculan indicadores de correlación o las probabilidades de emparejarse entre dos individuos del mismo nivel educativo, únicamente considerando al universo de parejas identificadas en cada fuente de información disponible. Sin embargo, el problema inherente a este enfoque radica en que no se tiene en cuenta que pueden existir distintos incentivos para emparejarse y que las características de las personas que optan por hacerlo pueden diferir de aquellas que eligen permanecer solas. Esto implica que el estudio del emparejamiento selectivo podría estar sesgado si no se consideraran las características de la población que decide emparejarse. Las consecuencias de esta omisión en la medición podrían ser análogas a los problemas que surgen al examinar los cambios en las distribuciones. En el caso de los cambios en las distribuciones, los indicadores podían estar computando cambios en los patrones de emparejamiento cuando no los había; en este caso, los indicadores podrían estar ignorando los cambios cuando sí los hubiera.

La evidencia presentada respalda una caída en el porcentaje de individuos en pareja en América Latina y el Caribe —fenómeno conocido como *unmating*— así como una selección negativa en términos educativos por parte de las mujeres a la hora de formar pareja. Dado

²⁰ 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.

este panorama, se considera necesario incorporar un indicador que no solo estime el patrón de emparejamiento entre quienes eligen hacerlo, sino que también tenga en cuenta qué características educativas poseen en relación a aquellos que no forman pareja. Para ello, se introduce el “Indicador Cuantílico Educativo Global” (ICEG) el cual se define de manera similar al ICEL con la única diferencia de que los cuantiles, en lugar de ser estimados en el universo de parejas, se estiman en función de la población total de individuos que potencialmente podrían emparejarse.

De esta manera, el valor del indicador reflejará la probabilidad de que dos personas pertenecientes a cada cuantil educativo se emparejen en relación con la probabilidad que surgiría en un escenario en el cual los patrones de emparejamiento fueran aleatorios, no solo en términos de la forma en que las parejas se formarían, sino también en términos de quiénes serían los individuos que conformarían dichas parejas.²¹

Indicador Geográfico

Los niveles de emparejamiento selectivo y sus tendencias han sido estudiados, en general, a nivel nacional.²² En países donde las áreas geográficas no son lo suficientemente extensas o donde las disparidades espaciales no revisten una significancia considerable, la omisión de la desagregación espacial podría no presentar una problemática relevante. Sin embargo, en el contexto de América Latina y el Caribe, este aspecto adquiere particular relevancia. La región latinoamericana se caracteriza por la presencia de países con dimensiones geográficas considerables, una tradición de baja movilidad a nivel nacional y marcadas desigualdades en términos espaciales en relación a los años de educación alcanzados (De La Mata et al., 2022). A título de ejemplo, en lo que respecta a las disparidades espaciales, en el censo correspondiente a la década de 2000 en México, la unidad subnacional de segundo nivel con el nivel educativo más alto registraba, en promedio, un nivel de educación 10 veces mayor que la de menor instrucción. Otra razón que justifica el interés en analizar el emparejamiento selectivo desde la perspectiva de las diferencias espaciales radica en la posibilidad de obtener una apreciación aproximada de los mercados matrimoniales locales en la región.

En los casos en los que el ES ha sido estudiado a nivel regional, se ha utilizado el coeficiente de correlación en años de educación entre las parejas en cada unidad subnacional.²³ Es pertinente resaltar que esta métrica podría ser objeto de una crítica similar a la que se formuló en relación al ICEL presentado en la sección anterior: al abordar el emparejamiento desde su perspectiva espacial, el coeficiente de correlación se limita a evaluar la relación lineal entre individuos pertenecientes a una misma región, sin considerar el nivel educativo promedio de la misma. En otras palabras, el coeficiente de correlación sería igual a 1 en caso de un emparejamiento perfecto, sin importar si la región considerada presentara un promedio educativo mayor o menor en el contexto de cada país. En este sentido, esta métrica no toma en cuenta que un mayor emparejamiento en regiones con diferentes promedios de años de educación puede tener consecuencias diversas en lo que respecta a la desigualdad, la movilidad o la pobreza a nivel nacional.

²¹ 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.

²² El CEDLAS presenta mapas de emparejamiento selectivo. Sin embargo, la desagregación espacial que hace es a un nivel considerablemente agregado en comparación a los datos que se presentarán en este caso.

²³ Este tipo de métricas es estimado, por ejemplo, por el CEDLAS para estudiar las diferencias regionales de emparejamiento.

En base a dicha crítica, se presentará una segunda variante del ICEL, denominada “Indicador Cuantílico Educativo Geográfico” (ICE Geo). Este indicador estudia la probabilidad de que dos individuos pertenecientes a una misma unidad geográfica y a un determinado cuantil educativo se junten, en comparación con la probabilidad en un escenario en el cual el emparejamiento es completamente aleatorio entre aquellos que optan por formar parejas, y donde la distribución en cada unidad administrativa es equivalente a la distribución a nivel nacional.²⁴ Este indicador, al igual que las versiones anteriores, puede calcularse de manera agregada o según cuantiles educativos específicos.

3.e. Funcionamiento del indicador

En esta sección se evaluará el desempeño del indicador propuesto en dos dimensiones clave. La primera se relaciona con la capacidad de este para medir patrones de emparejamiento controlando por cambios en las distribuciones marginales de educación. La segunda se enfoca en la relación entre los resultados obtenidos a través del ICEL con aquellos que surgen de otros indicadores reconocidos como los más robustos para medir este fenómeno, como los coeficientes de correlación de rangos.

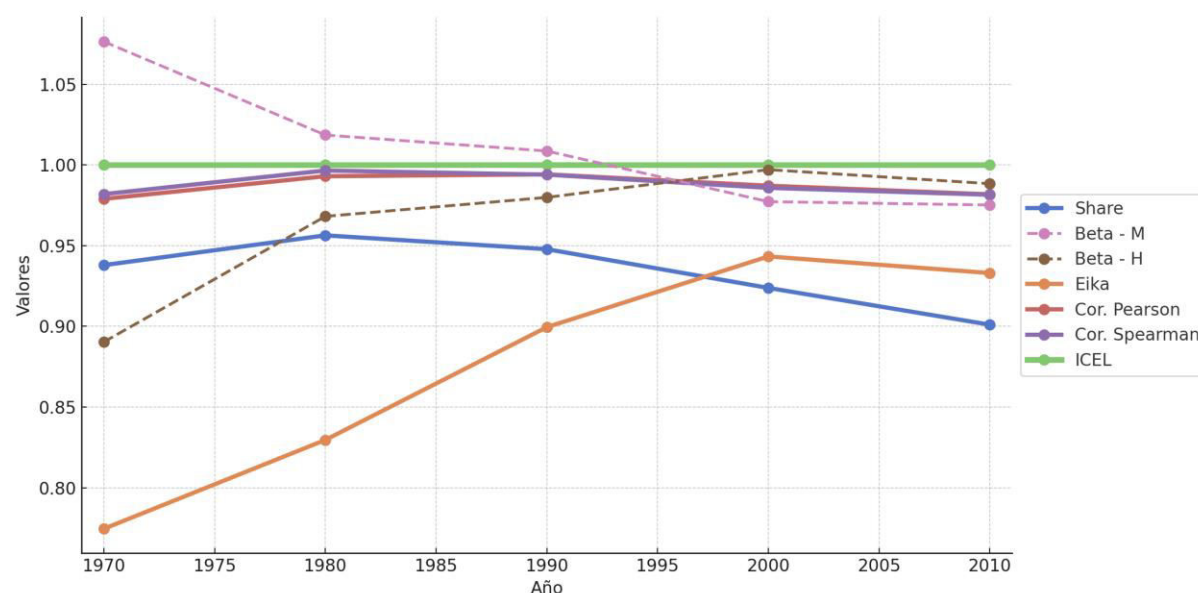
Para analizar la capacidad de medir los patrones de emparejamiento controlando por las variaciones en las distribuciones educativas, se utilizarán datos de Argentina. En cada año, se simulará un escenario de emparejamiento selectivo perfecto en términos educativos, donde la mujer con mayor nivel educativo se emparejará con el hombre más educado, la segunda mujer lo hará con el segundo hombre más educado, y así sucesivamente hasta completar todas las parejas. Si bien el patrón de emparejamiento, por construcción, se mantendrá constante en la simulación, las distribuciones educativas de hombres y mujeres pueden cambiar año a año según las variaciones que se registraron en la realidad. De esta forma, cualquier variación en el patrón de emparejamiento solo se producirá como resultado de cambios mecánicos en las distribuciones educativas. En este contexto, las métricas de emparejamiento selectivo que controlen efectivamente por cambios en las distribuciones no deberían mostrar modificaciones a lo largo del tiempo.

En este ejercicio de microsimulación se incluirá el indicador propuesto, así como también varios de los indicadores presentados en la revisión previa. Entre aquellos de la primera generación se encuentran el *share* de individuos con el mismo nivel educativo y los coeficientes de regresión (alternando la variable dependiente e independiente de la regresión). Entre los de segunda generación, se incluye el indicador de Eika. Desafortunadamente, dado que el ejercicio de simulación de emparejamiento perfecto resulta en muchas celdas sin emparejamientos—por ejemplo, nunca hay emparejamientos entre un individuo con primaria incompleta y un graduado universitario en la simulación de emparejamiento perfecto—no se pueden medir las tendencias utilizando el modelo log-lineal empleado por Schwartz y Mare (2005). Por último, entre los indicadores de tercera generación incluyen a los indicadores de correlación lineal de Pearson y al de correlación lineal de Spearman, así como también el ICEL.

²⁴ Mientras el rango del indicador en la sección anterior estaba entre 0 y el número de cuantiles, en este caso, estará entre 0 y el número de cuantiles al cuadrado. Por ejemplo, si se consideraban quintiles, el mayor emparejamiento estaba dado por $0.2/(0.2*0.2) = 5$. En este caso, el máximo valor puede estar dado por $1/(0.2*0.2) = 25$. Dicho caso se daría cuando una unidad administrativa está formada únicamente por individuos de un solo cuantil. En ese caso, el 100% de hombres y mujeres emparejados pertenecerían al mismo cuantil, y el contrafáctico sería 0.04.

En la Figura 3, se presentan los resultados de cada indicador en cada uno de los escenarios simulados. Vale aclarar que la métrica educativa sobre la cual mide ES cada indicador puede variar —años de educación, categorías educativas, cuantiles de una distribución educativa— así como también que los indicadores son normalizados entre 0 y 1, en caso de que esto sea necesario.

Figura 3. Resultados de cada indicador bajo ES perfecto.



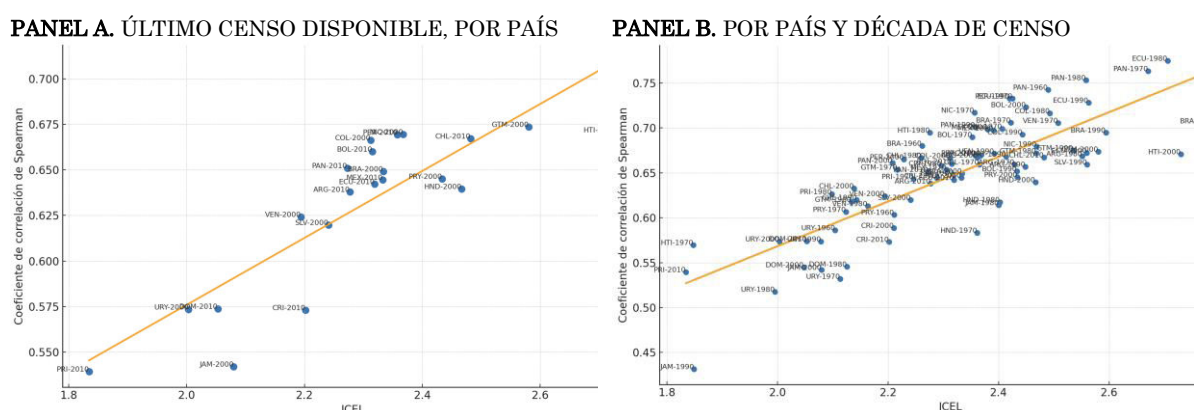
Nota: La figura muestra la evolución del emparejamiento selectivo en un escenario simulado de ES perfecto, a través de la estimación de diversos indicadores. Estos incluyen: el porcentaje de parejas con el mismo nivel educativo (*share*), los coeficientes de regresión donde la educación de la mujer es 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).

Por construcción, el ES es perfecto en todos los escenarios; por lo tanto, cuanto más cercanos a 1 sean los resultados obtenidos, mejor será el desempeño de cada indicador. En esa línea, los resultados de la figura dejan en evidencia la superioridad que representan los indicadores de tercera generación —los coeficientes de correlación de rangos y el ICEL— para medir el ES controlando por cambios en las distribuciones educativas.

La segunda dimensión a considerar se refiere a que el ICEL se enfoca exclusivamente en los patrones de emparejamiento entre individuos que pertenecen a la misma categoría educativa. Por lo tanto, si el emparejamiento agregado dependiera en gran medida de interacciones entre distintas categorías educativas, más allá de aquellas con el mismo nivel, el ICEL podría ofrecer una perspectiva limitada al no contemplar la distribución completa. Para abordar esta potencial limitación, se evalúa el desempeño del indicador propuesto en comparación con el coeficiente de correlación de rangos de Spearman. En la Figura 4 se presenta un gráfico de dispersión que compara las métricas del ICEL con las del coeficiente de Spearman, tanto para el último dato disponible en cada país (Panel A) como para todos los períodos analizados (Panel B).

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 naranja 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).

La relación entre los valores de ambas métricas es significativamente alta. Para el último dato disponible, se observa una correlación de 0.85 entre el ICEL y el coeficiente de Spearman. Los resultados, a su vez, subrayan una de las virtudes del indicador: su capacidad para generar resultados claros y fácilmente interpretables. En este marco, un valor de correlación cercano a 0.675 indica que, en promedio, la probabilidad de que dos personas se emparejen con individuos pertenecientes a su mismo quintil educativo es 2.5 veces mayor en comparación a la que surgiría en un escenario de emparejamiento aleatorio.

En resumen, esta subsección deja en evidencia la eficacia de los indicadores de tercera generación para medir la evolución del ES controlando por los cambios en las distribuciones. Además, destaca que el ICEL es útil para capturar tendencias a nivel agregado, a pesar de que se enfoque exclusivamente en individuos con el mismo nivel educativo.

4. Resultados

Una vez identificados el ICEL y el coeficiente de correlación de rangos de Spearman como indicadores robustos para evaluar el nivel y la evolución del emparejamiento selectivo en la región, se emplearán ambos para analizar los niveles de ES en los distintos países de América Latina y el Caribe, así como su evolución a lo largo del tiempo. Además, se presentarán estimaciones del ES por nivel educativo y se aplicarán las extensiones metodológicas propuestas.

4.a. Emparejamiento Selectivo en la región

En la Figura 5, se presentan los resultados del Indicador Cuantílico Educativo Local Agregado, los valores del ICEL por quintil y los coeficientes de correlación de rangos de Spearman para los últimos registros disponibles de cada país.

Los resultados de ambos indicadores –ICEL agregado y Spearman— presentan evidencia a favor de emparejamiento selectivo positivo para todos los países de la muestra. Asimismo, se observa que el grado de emparejamiento selectivo varía tanto entre los diferentes países de la muestra como entre los quintiles dentro de cada país. Por ejemplo, mientras que el promedio simple regional muestra que los individuos son 2.3 veces más propensos a emparejarse con alguien de su mismo quintil educativo en comparación con un emparejamiento aleatorio, se registran valores cercanos a 2 en países como Puerto Rico, Uruguay y República Dominicana, y valores cercanos a 3 en Guatemala y Haití. Además, la media regional del indicador para el quintil 1 es 2.5, con un rango que oscila entre 2 y 3.5, mientras que para el quintil 5, la media es de 2.97, con valores que van de 1.65 a 3.5. Por último, se puede observar que el ordenamiento de los valores del coeficiente de correlación de rangos de Spearman –que asciende en promedio a 0.63—, en términos generales, es consistente con los resultados obtenidos a través del ICEL.

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



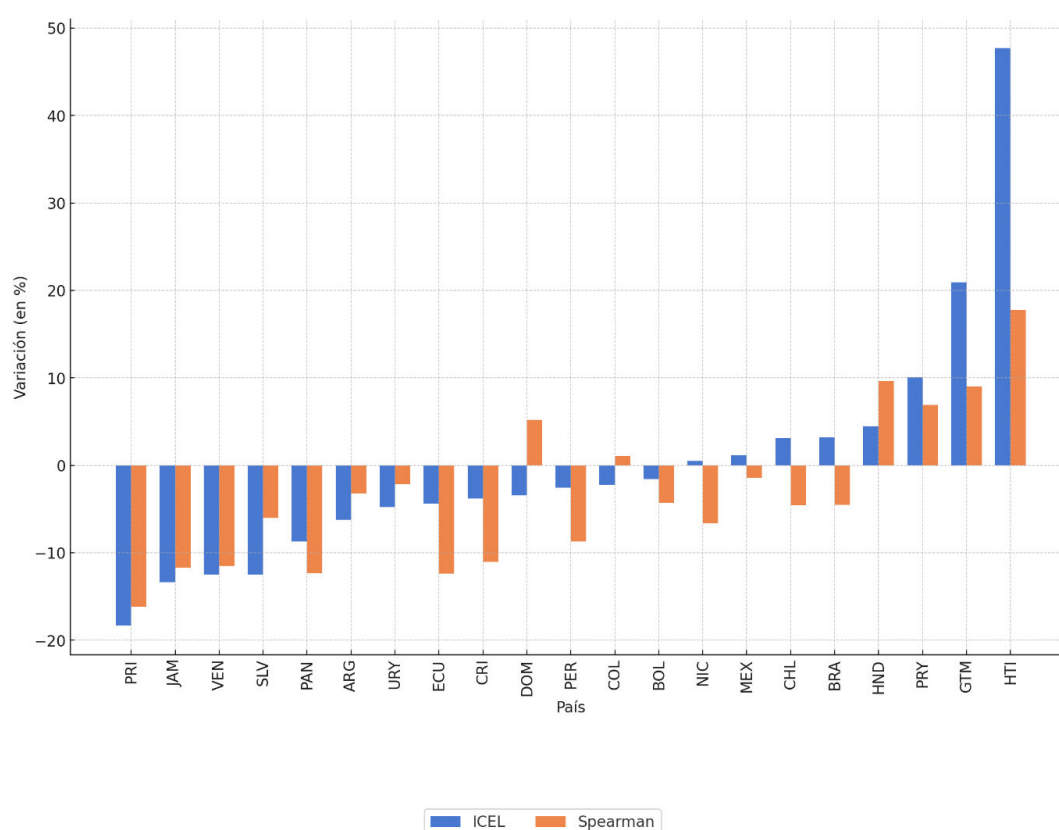
Nota: La figura muestra el valor agregado del coeficiente de Spearman y del ICEL agregado junto con los valores del ICEL para cada uno de los cuantiles de la distribución educativa. Los quintiles 2 a 4 se representan en color gris. Los valores del ICEL se ubican en el eje inferior, mientras que en el 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.b. Evolución del ES en la región

Como se señaló en las secciones anteriores, es relevante analizar la evolución del fenómeno en cada país de la región para comprender si los cambios en la formación de parejas podrían estar vinculados con variaciones en la desigualdad o la movilidad. En la Figura 6, se presenta la variación porcentual del ICEL y del coeficiente de Spearman entre el primer y el último registro disponible en cada país (en la [Figura A3](#), se presenta la evolución de los indicadores para cada uno de los años disponibles, por país). Los resultados muestran que, según el ICEL, 13 de los 21 países presentan una caída en los niveles de emparejamiento selectivo entre períodos. Por su parte, los valores del coeficiente de Spearman reflejan caídas más pronunciadas en 15 de los 21 países.

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)

En algunos casos —como Puerto Rico, Jamaica, Venezuela, El Salvador y Panamá— se registran caídas cercanas al 10% en base a cualquiera de los indicadores. En contraste, países como Honduras, Paraguay, Guatemala y Haití muestran aumentos significativos en los niveles de ES a lo largo del tiempo. Adicionalmente, se identifica un grupo de países donde la magnitud de la disminución varía según el indicador utilizado (Argentina, Uruguay, Ecuador, Costa Rica y Perú), aunque las caídas suelen ser moderadas. Por último, algunos países prácticamente no muestran variaciones, con diferencias que varían dependiendo del indicador considerado.

Al comparar las estimaciones de este análisis con las estadísticas de SEDLAC (CEDLAS y Banco Mundial, 2024), una base de datos que armoniza indicadores distributivos, laborales y sociales en la región en base al procesamiento de encuestas de hogares, se observan diferencias en los niveles de los indicadores para ciertos países, aunque las tendencias generales se mantienen consistentes ([Figura A4](#))²⁵. Este análisis permite observar la evolución del fenómeno desde aproximadamente la década de 1960 hasta 2022,

²⁵ En este caso, se compara la evolución del coeficiente de correlación lineal de Pearson, también identificado como uno de los indicadores robustos para medir la evolución del ES en la sección 3.

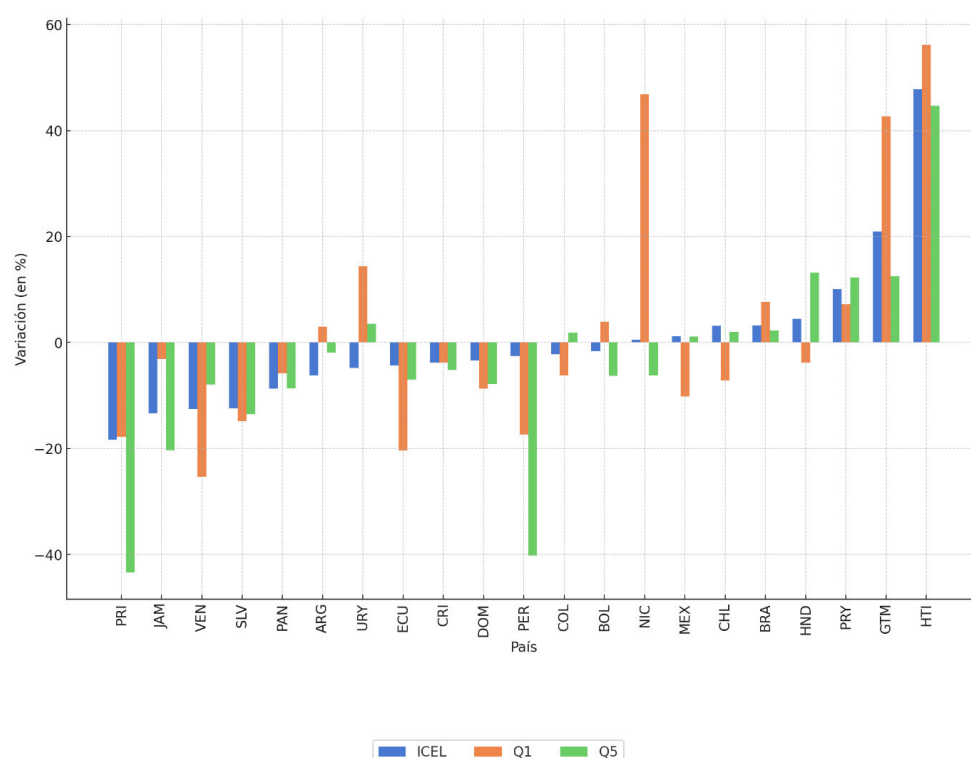
concluyendo que la región ha experimentado una disminución generalizada en los niveles de emparejamiento selectivo

4.c. Evolución del ES por quintiles.

Uno de los puntos de interés en el estudio de los patrones de ES se refiere a la evolución del mismo en las distintas categorías de la distribución educativa, más allá del nivel agregado. Es posible que la disminución generalizada del emparejamiento selectivo a nivel agregado oculte tendencias heterogéneas según el quintil educativo, lo que implicaría consecuencias, potencialmente, diferentes sobre la desigualdad. En la Figura 7 se presenta la evolución del ICEL tanto a nivel agregado como para los dos extremos de la distribución, los quintiles Q1 y Q5, entre el primer y el último registro disponible de cada país (en la [Figura A5](#), se presenta la evolución de los indicadores para cada uno de los años disponibles, por país).

En general, se observa que las tendencias entre el nivel agregado y los extremos de la distribución son similares, siendo excepcionales los casos en los que se registran patrones opuestos. Algunos ejemplos de resultados contrapuestos incluyen a Argentina, Uruguay y Bolivia, casos en los cuales, si bien el indicador agregado disminuye en el tiempo, se observa un aumento del ES entre los menos educados. Por otro lado, en Honduras se evidencia un comportamiento inverso al de estos países: un aumento del ES a nivel agregado, acompañado por una disminución del ES entre los menos educados. En el caso de Nicaragua, se observa un aumento pronunciado del ES en el quintil más bajo, mientras que en el quintil superior se registra una disminución. En consecuencia, el promedio agregado del país refleja un nivel de ES prácticamente estable.

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)

En línea con lo señalado, al comparar las tendencias por quintiles con las agregadas, se observa que los extremos de la distribución parecen ser los principales determinantes del comportamiento agregado del indicador y de los coeficientes de correlación de rangos. Por ejemplo, la correlación entre el ICEL para el quintil superior (Q5) y el coeficiente de correlación de rangos de Spearman alcanza un valor de 0.8, mientras que para el quintil inferior (Q1) asciende a 0.6.

4.d. Indicador Cuantílico Educativo Global

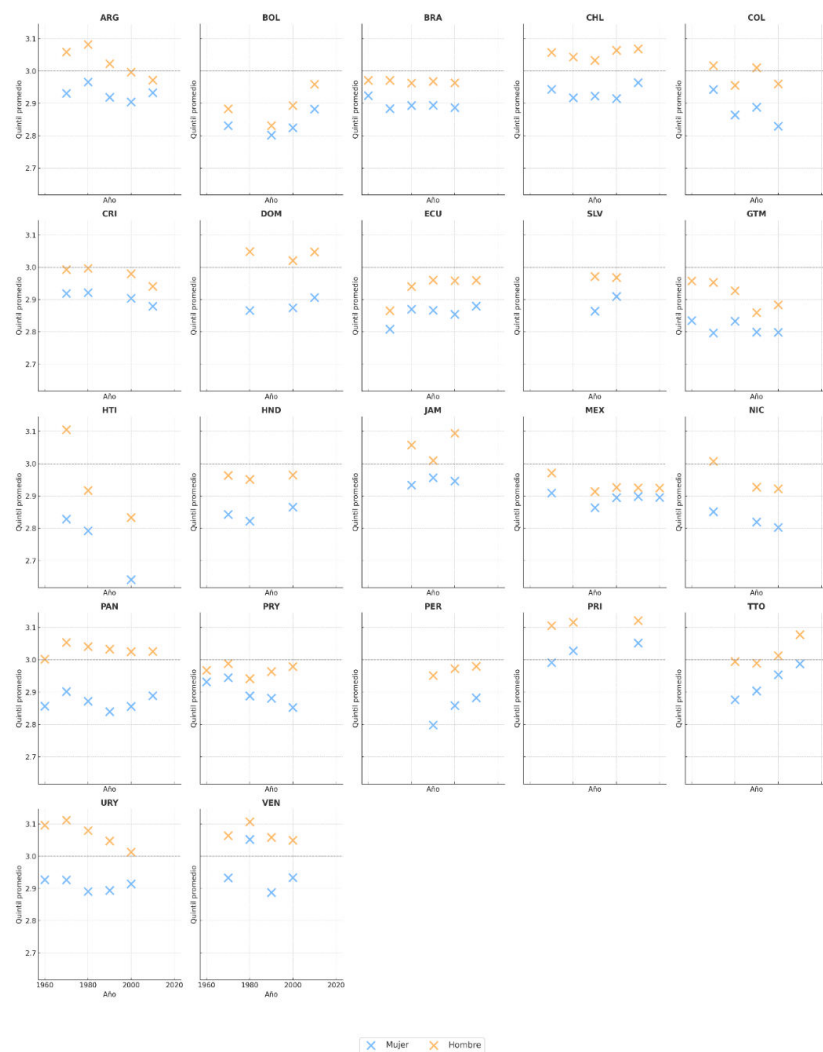
Una de las extensiones del ICEL radica en la capacidad de estudiar el emparejamiento considerando al universo de personas que potencialmente podrían formar una pareja, en lugar de restringir el análisis únicamente a los individuos que deciden emparejarse. Para entender la relevancia de esta extensión, vale recalcar que en la región ha habido importantes cambios en los patrones de formación de parejas a lo largo del tiempo. 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.

En línea con dicha caída, podría suceder que las características de la muestra más restringida que decide formar parejas difieran de aquellas lo hacían en el pasado. En la

Figura 8, se presenta el quintil promedio de la distribución educativa del total de individuos por cohorte de nacimiento y género, al que pertenecen los individuos en pareja. Si no existiera ningún tipo de selección a la hora de emparejarse, el quintil promedio de estos grupos debería ser igual a 3. En caso contrario, la evidencia respaldaría un patrón de selección positiva o negativa. Los resultados muestran que, para todos los países y para todos los años bajo estudio —con excepción de Puerto Rico—, se observa que las mujeres se ubican por debajo del quintil 3, mientras que los hombres siempre se ubican por encima de ellas, aunque no necesariamente por encima del quintil 3. En términos de tendencias, no se observan patrones sistemáticos a lo largo del tiempo, a pesar de la caída en la proporción de individuos en pareja.

La Figura 9 compara los resultados del ICEL con el ICEG para los últimos registros disponibles (ver [Figura A7](#) para estudiar la evolución de cada indicador). En general, no se observan grandes diferencias entre los indicadores, pero se puede evidenciar que, en la mayoría de los países, el indicador que incluye la selección de quienes deciden emparejarse (global) tiende a presentar valores mayores a aquel que considera únicamente el universo de individuos en pareja.

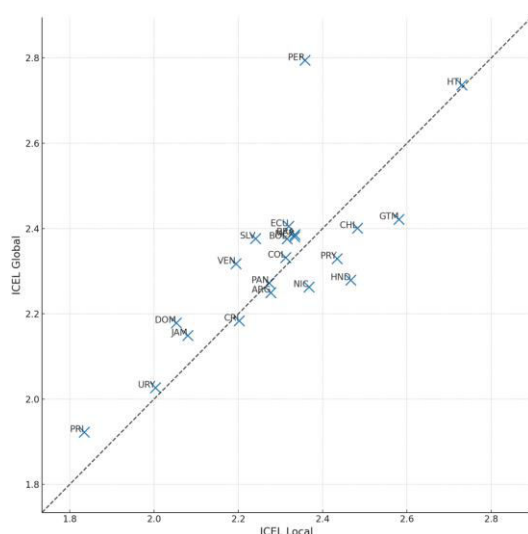
Figura 8. Quintil promedio de los individuos en pareja, por país y década



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).

Figura 9. Comparación entre los niveles del ICEG e ICEL



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

Fuente: elaboración propia en base a IPUMS (2020)

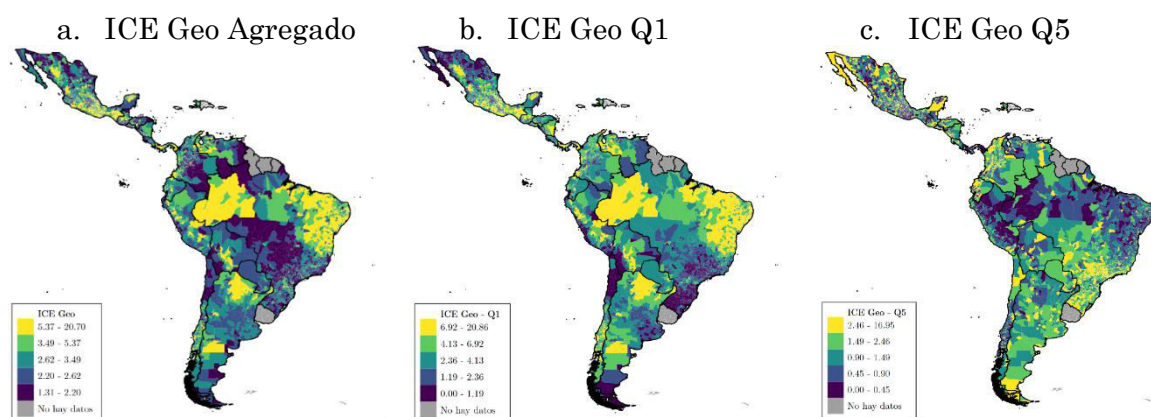
4.e. Indicador Cuantílico Educativo Geográfico

Como se comentó en la Sección 3, la literatura enfocada en estudiar los niveles de emparejamiento a nivel regional, en general, se ha basado en coeficientes de correlación por provincia o departamento de cada país. Si uno aplicara este análisis para la región, podría observar que los coeficientes de correlación evidencian una notable heterogeneidad en los niveles de emparejamiento selectivo dentro de cada uno de los países, con algunas regiones en las cuales el ES resulta particularmente alto: el sur de México, el sudeste de Brasil, el centro de Chile, el centro y norte de Argentina, el sudoeste de Bolivia y Guatemala (ver [Figura A8](#) en el Apéndice). Sin embargo, este tipo de indicadores omiten el nivel educativo promedio de la región. En ese sentido, dos unidades subnacionales pueden arrojar los mismos coeficientes de emparejamiento, cuando los niveles educativos promedio de cada una difieran sustancialmente entre ellas, y consecuentemente, las implicancias de ese emparejamiento serían diferentes en cada caso.

En la Figura 10, se presentan los resultados de la segunda posible extensión del indicador de emparejamiento correspondientes al censo de la década del año 2000.²⁶ Estos tienen como propósito complementar el análisis de correlaciones por región, combinando los patrones de emparejamiento, que ya se mostró que presentan una gran heterogeneidad, con los niveles promedio de educación por región. En el Panel A, se presentan las estimaciones del Indicador Cuantílico Educativo Geográfico, mientras que en los Paneles B y C se exponen los resultados del Indicador por categorías correspondientes al primer y quinto quintil, respectivamente, de la distribución educativa de cada país.

²⁶ Se eligió esta década ya que es la que mayor cobertura tiene en términos de países disponibles.

Figura 10. Indicador de Bratsberg espacial – América Latina y el Caribe, 2000



Nota: el gráfico presenta los valores que toma el ICEL por unidad subnacional de segundo orden en la década de los 2000. En el Panel A, se presentan los valores del ICEL 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)

Las figuras presentan hallazgos novedosos en relación con la manera en que las personas establecen relaciones en la región y entre quiénes lo hacen.

En primer lugar, al utilizar el ICE Geográfico agregado, se evidencia que los niveles de emparejamiento muestran notables disparidades a nivel geográfico, dentro de cada país. Al comparar estos resultados con los coeficientes de correlación por unidad administrativa, se observa que, en algunos casos, las regiones con un mayor emparejamiento se mantienen (por ejemplo, en el centro y norte de Argentina, el centro de Chile, el sur de México, Guatemala y el suroeste de Bolivia), mientras que, en otros casos, los patrones varían (por ejemplo, en Brasil). En el agregado, sin embargo, la correlación entre métricas es nula.

En segundo lugar, los paneles B y C, presentan resultados naturalmente contrapuestos: generalmente, en las áreas donde es más probable que las personas altamente educadas se emparejen son justamente aquellas donde es menos probable que las menos educadas lo hagan –la correlación es igual a -0.53.

Las diez unidades geográficas de América Latina y el Caribe con mayores probabilidades de que emparejamiento dos individuos del quintil superior de la distribución educativa: Antiguo Cuscatlán, Nuevo Cuscatlán y Santa Tecla (en La Libertad, El Salvador); Vitacura, Providencia, Las Condes y Ñuñoa (en Santiago, Chile) y San Pedro Garza García (en Nuevo León, México), Asunción (Paraguay), San Andrés Huayapam (México) y Envigado (Colombia). Estos valores surgen como una combinación del nivel educativo promedio y de los patrones de emparejamiento de las unidades geográficas, lo cual puede estar directamente vinculado con altos niveles de segregación y desigualdad regional en cada uno de esos países. En Argentina, la mayor probabilidad se encuentra en la ciudad capital de Mendoza, Vicente López, la Ciudad Autónoma de Buenos Aires, San Isidro y la capital de San Juan. Por otro lado, las probabilidades de que se emparejen dos individuos con menos educación aumentan significativamente en municipios de provincias del norte del país, en Misiones, Chaco, Salta y Santiago del Estero. Si bien los valores del indicador, naturalmente, están afectados por el nivel educativo de cada región, la relación no es perfecta. Por ejemplo, la correlación entre años de educación y ICE Geográfico para individuos del Quintil 5 es igual a 0.625 (y a -0.7 para aquellos del Quintil 1).

A pesar de que los datos se refieren a censos realizados en la década de 2000, la distribución del emparejamiento ha mantenido cierta estabilidad a lo largo del tiempo. Además, aunque la información se presenta a nivel de segunda agregación geográfica, los patrones se mantienen al considerar la desagregación administrativa de primer nivel. No obstante, al modificar la unidad administrativa, los patrones pueden variar en cuanto a nivel y mostrar una mayor variabilidad.

En resumen, los resultados que consideran la dimensión geográfica muestran que existen grandes diferencias espaciales en los niveles educativos dentro de cada país de América Latina y el Caribe, y que dichas diferencias, junto con patrones de emparejamiento diferenciales, se traducen en diferentes probabilidades que tienen los individuos más y menos educados del país de emparejarse. Con base en estos resultados, se puede concluir que la dimensión geográfica emerge como un factor de considerable relevancia en la comprensión de los patrones de emparejamiento selectivo y sus consecuencias.

5. Determinantes del emparejamiento

Simulaciones de emparejamiento aleatorio

El objetivo de esta sección es determinar la influencia que pueden ejercer diferentes dimensiones sobre los patrones de emparejamiento selectivo a nivel agregado. En particular, motivados por la heterogeneidad documentada en términos geográficos en la sección previa, se pretende estimar la importancia de la dimensión geográfica, la edad y el estado de emparejamiento en el nivel de ES agregado de un país.

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 emplearán datos correspondientes a los años 2000 y 2020 para México y, para este ejercicio en particular, se ampliará la muestra a individuos de entre 26 y 60 años. A continuación, se describen los seis escenarios alternativos objeto de simulación:

1. En el primer escenario, se simulará un patrón de emparejamiento completamente aleatorio en términos educativos en el que cualquier individuo, más allá del su estado de emparejamiento, pueda formar una pareja. Se restringirá el número de nuevas parejas al observado en la realidad.
2. El segundo escenario mantendrá constante el estado de formación de pareja de cada individuo, lo que significa que las parejas generadas en cada simulación estarán compuestas exclusivamente por personas que, en el escenario real, se encontraban efectivamente cohabitando.
3. El tercer escenario sumará una restricción adicional al escenario anterior. En este, además del estado de emparejamiento, se mantendrá invariable la distribución de rangos etarios de la población, aleatorizando a los individuos de cada cohorte por separado. Los rangos etarios se definirán de la siguiente manera: de 26 a 29 años; de 30 a 39; de 40 a 49; y de 50 a 60. Por ejemplo, una pareja real conformada por un hombre de entre 30 y 39 años y una mujer de entre 26 y 29 años se reemplazará por dos individuos del mismo rango etario, según el género.
4. El cuarto escenario tendrá en cuenta el factor geográfico, de modo que, a las restricciones previas, se sumará una nueva: la aleatorización se realizará entre individuos que pertenecen a la misma unidad administrativa de *primer* orden.

5. El quinto escenario plantea la misma restricción que el cuarto, pero con la diferencia que la restricción geográfica es aún más fuerte: las nuevas parejas deberán estarán por individuos que pertenecen a la misma unidad geográfica de *segundo* orden.

El emparejamiento selectivo se medirá empleando siete métricas para los diferentes escenarios: el coeficiente de correlación de rangos de Spearman y los Indicadores Cuantílicos Educativos Local y Global, tanto en su versión agregada como también las correspondientes a los quintiles extremos de la distribución.

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, los datos indican que el grado de emparejamiento selectivo, según todos los indicadores agregados, se redujo entre el año 2000 y el 2020 en México. En segundo lugar, vinculado con el funcionamiento del indicador propuesto, al examinar detenidamente los resultados de las dos primeras aleatorizaciones, se hace evidente la diferencia entre el indicador global y aquellos que consideran únicamente el universo de parejas existentes en la estimación. Mientras que los coeficientes de correlación y las métricas del ICEL no presentan variaciones entre la primera y segunda simulación (R1 y R2), lo que sugiere una completa aleatorización en ambos escenarios, el ICEG presenta variaciones entre escenarios. Dado que este último considera a la población de potenciales individuos en pareja, señala, por ejemplo, que en México, en un escenario de emparejamiento aleatorio entre los individuos que conforman parejas, las probabilidades de que dos individuos del quinto quintil educativo –definidos en base a la distribución educativa de toda la población– se emparejen son menores que las que surgirían si el emparejamiento fuera aleatorio tanto entre quienes se emparejan como entre los individuos que toman la decisión de formar parejas.

En tercer lugar, al considerar la restricción etaria, se ve un aumento de entre 4 y 9 puntos en el coeficiente de correlación de Spearman, que surge justamente porque las edades

estarán vinculadas con el nivel educativo promedio de las personas que se están emparejando. Por otra parte, los aumentos en el ICE (en sus dos versiones) es menor, justamente, porque su definición se basa en cuantiles de la distribución educativa para cada una de las edades.

Finalmente, cuando se añaden restricciones geográficas, estas parecen desempeñar un papel de gran relevancia al explicar los patrones de emparejamiento. Por un lado, cuando se consideran las unidades administrativas de primer orden, se advierte que la sumatoria de las tres restricciones —estar en pareja, grupo etario y *geolevel 1*— explica aproximadamente una quinta parte de la correlación educativa observada, y la probabilidad de encontrar a individuos pertenecientes al mismo quintil asciende a valores cercanos a 1,20 en los dos censos. Cuando se vuelve más restrictiva la dimensión geográfica, en la simulación 5, se constata que la correlación educativa que surge de un escenario aleatorio llega a explicar más de un tercio de la correlación observada en cada período.²⁷

Mientras los resultados de la sección previa pusieron de manifiesto que en América Latina y el Caribe existe una significativa disparidad espacial en lo que respecta tanto a los años de educación alcanzados, como a los patrones de emparejamiento y, por consiguiente, a las probabilidades de que dos individuos que pertenecen a los quintiles más y menos educados formen una pareja en una determinada unidad administrativa subnacional. Motivados por dicha heterogeneidad, estudiamos el peso relativo de la geografía en los patrones de emparejamiento y observamos que las disparidades espaciales en un escenario simulado de emparejamiento aleatorio pueden explicar hasta un sexto del emparejamiento total observado en la práctica. En resumen, la dimensión geográfica parece emerger como un factor determinante a la hora de explicar los patrones de emparejamiento.

6. Emparejamiento selectivo y desigualdad

La literatura sobre emparejamiento selectivo se ha preguntado, de manera recurrente, cómo se vería afectada la desigualdad de ingresos en los hogares si las parejas se formaran de manera aleatoria en lugar de seguir los patrones observados en la realidad (Eika et al., 2019; Greenwood et al., 2014; Hryshko et al., 2017, para Estados Unidos; Leal, 2015; Pereira y Santos, 2017 para países latinoamericanos).

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

²⁷ Un punto relevante para considerar en esta simulación es que las unidades administrativas de segundo orden en México pueden ser demasiado pequeñas como para suponer que los individuos podrían emparejarse únicamente con personas de su misma unidad.

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.

En esta subsección, se estudiará la relación entre emparejamiento selectivo y desigualdad considerando el enfoque aditivo. En particular, se simularán los mismos escenarios que en la Sección 5. Se considerará el ingreso total laboral de las parejas, definido como el ingreso proveniente de sueldos, salarios, comisiones o venta de cultivos, animales. No se ajustará el ingreso por número de integrantes en el hogar y se considerará únicamente el universo de parejas. En cada una de las simulaciones, la desigualdad en la distribución del ingreso se medirá a partir de tres indicadores alternativos: el coeficiente de Gini, la relación de ingresos entre el decil 10 y el decil 1, y el índice de Atkinson (1). En la Tabla 4, se presentan los resultados de cada una de las simulaciones.

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 simuladores. 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 resultados muestran cómo los patrones de emparejamiento selectivo podrían explicar parte de la desigualdad de ingresos y cómo esta contribución ha cambiado a lo largo del tiempo²⁸. En ambos años, el nivel de desigualdad observada es significativamente superior al que se obtendría bajo escenarios de emparejamiento aleatorio. Por ejemplo, en el año 2000, el coeficiente de Gini observado es de 56.4, mientras que en el escenario de aleatorización total (R1) disminuye a 53.1, lo que implica una reducción del 5.76%. De manera similar, en 2020, el Gini desciende de 46.5 a 43.5 bajo el mismo escenario, representando una reducción del 6.45%. Las variaciones en el ratio de ingreso entre deciles y en el Atkinson muestran patrones similares, con caídas de entre 15.1% y 17.1% en el primer caso, y cercanas al 9% en el segundo caso. Sin embargo, a medida que se introducen restricciones—como la formación de parejas, la consideración de grupos etarios y factores geográficos— el efecto de la aleatorización sobre los indicadores de desigualdad se atenúa en ambos años. En 2000 y 2020, el coeficiente de Gini se reduce progresivamente menos a medida que se incorporan estos factores. Por ejemplo, en el escenario R5 (que incluye pareja, grupo etario y factores geográficos), la disminución del Gini es de solo -2.82% en 2000 y -3.41% en 2020, en comparación con la aleatorización total (R1). Nuevamente, los resultados se repiten al considerar los otros indicadores de desigualdad.

²⁸ No se realizan evaluaciones de los cambios en la desigualdad entre los dos momentos por la potencial incompatibilidad de los conceptos de ingreso incluidos en cada año.

En suma, los resultados indican que si bien las simulaciones de ES bajo un escenario de emparejamiento aleatorio pueden dar una idea de cuánto este contribuye a moldear la desigualdad, al considerar ciertas restricciones que están naturalmente vinculadas a los patrones de formación de pareja, la caída en la desigualdad es moderada.

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 a 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 alto 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 espaciales 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-Orsorio, 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>
- 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. (2022). *Socio-Economic Database for Latin America and the Caribbean*.
<https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/sedlac/estadisticas/>
- CEDLAS y Banco Mundial. (2021). *SEDLAC* [Base de datos].
<https://www.cedlas.econo.unlp.edu.ar/wp/en/estadisticas/sedlac/>
- 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., Arreaza, A., & Alves, G. (2022). *Desigualdades heredadas. El rol de las habilidades, el empleo y la riqueza en las oportunidades de las nuevas generaciones*. Caracas.
<https://cafsciotea.azurewebsites.net/handle/123456789/1981>
- Eika, L., Mogstad, M., & Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6), 2795–2835.
- 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>
- Ganguli, I., Hausmann, R., & Viarengo, M. (2014). Marriage, education and assortative mating in Latin America. *Applied Economics Letters*, 21(12), 806–811.
- 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*.
- 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.
- 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.

Apéndice

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. Máximo valor que toma la variable “años de educación” en la base IPUMS.

País	Década 1960	Década 1970	Década 1980	Década 1990	Década 2000	Década 2010
Argentina	-	18	18	18	17	18
Bolivia	-	18	-	18	17	18
Brasil	17	16	17	17	17	-
Chile	18*	18	18	18	18	18
Colombia	17*	18	18	12	17	-
Costa Rica	18*	18	18	-	18	18
Dom. Rep.	18*	18*	17	-	18	18
Ecuador	18*	18	18	18	18	18
El Salvador	-	-	-	17	18	-
Guatemala	18	18	18	18	18	-
Haití	-	18	17	-	18	-
Honduras	12*	18	18	-	18	-
Jamaica	-	-	11	18	18	-
México	18*	18	-	18	18	18
Nicaragua	-	17	-	17	18	-
Panamá	18	18	18	18	18	18
Paraguay	18	18	18	18	18	-
Perú	-	-	-	12	12	-
Puerto Rico	-	18	18	-	-	18
Trinidad y Tobago	-	-	18	18	18	18
Uruguay	17	17	17	17	18	-
Venezuela	-	18	18	18	18	-

Nota: cada celda presenta el máximo valor que toma la variable años de educación en cada uno de los censos considerados. Los años marcados con * son censos en los cuales no es posible vincular a los habitantes del hogar.

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

Tabla A3 – Estadísticas descriptivas período completo

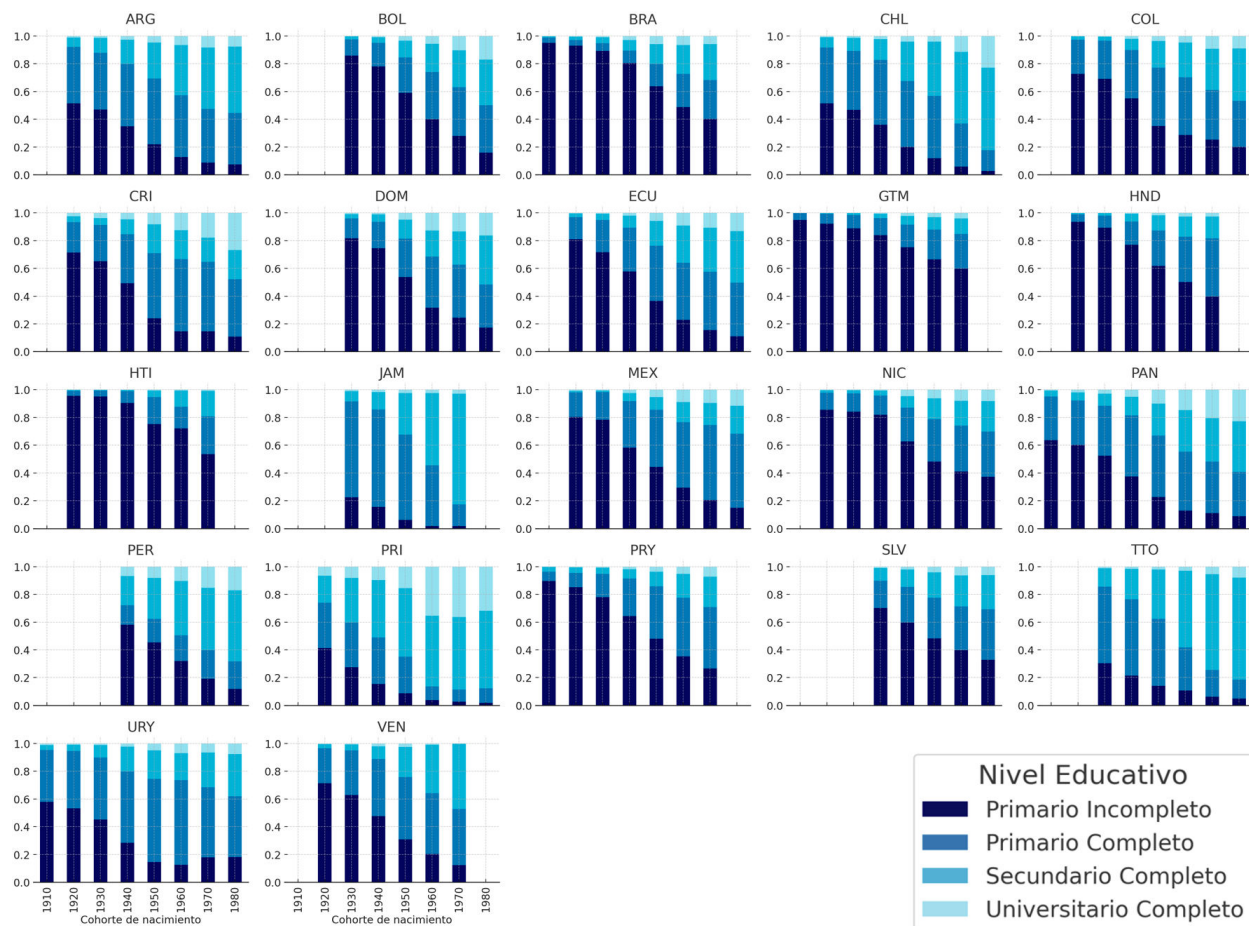
Panel A. En pareja												
Medias muestrales	1960		1970		1980		1990		2000		2010	
	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres
Share (%)	64.92	64.06	63.86	65.63	61.02	63.07	60.13	61.31	58.08	59.07	52.88	51.42
Edad	34.16	35.13	34.21	35.10	33.95	34.77	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.

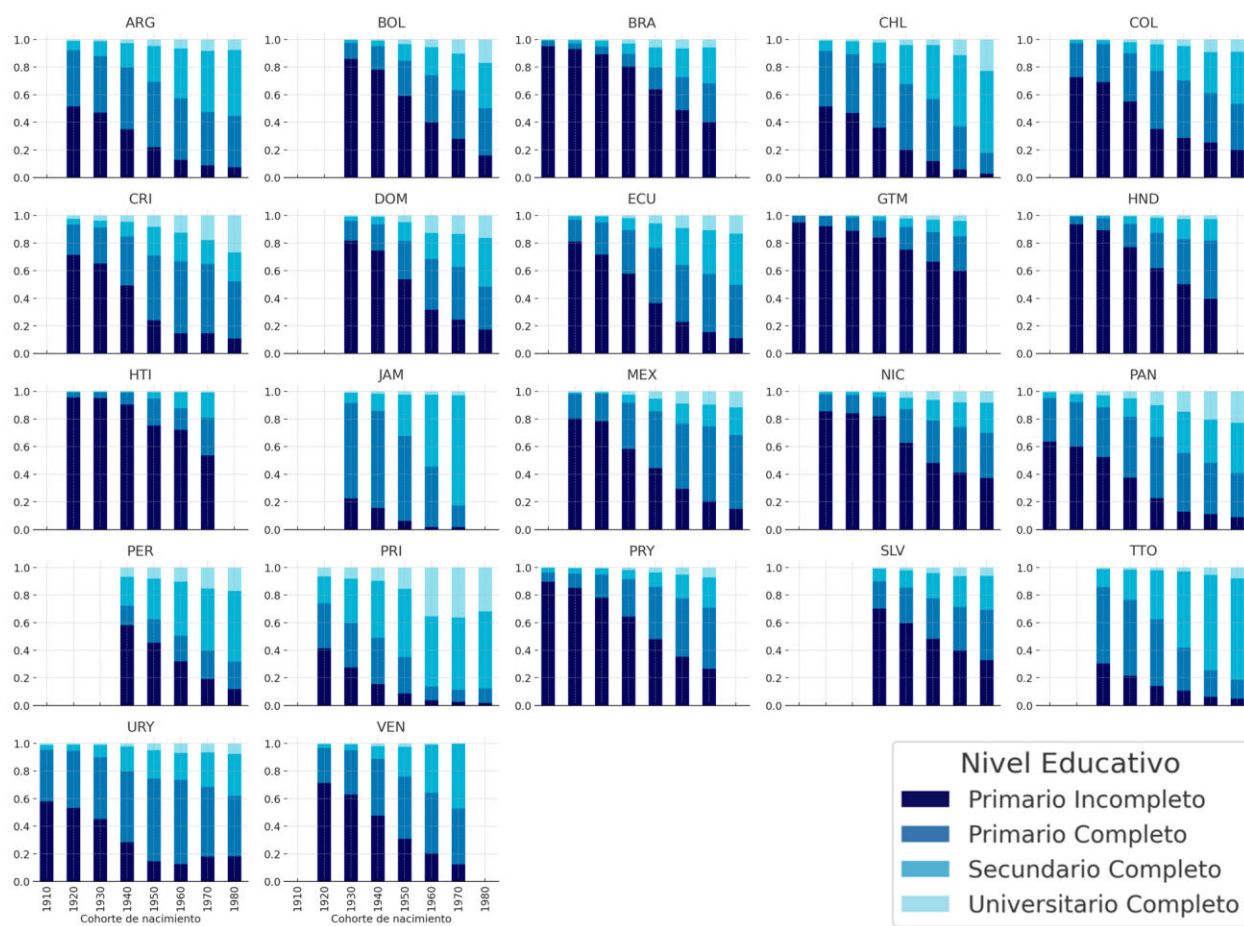
Fuente: IPUMS-I (2020)

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

Panel A. Hombres



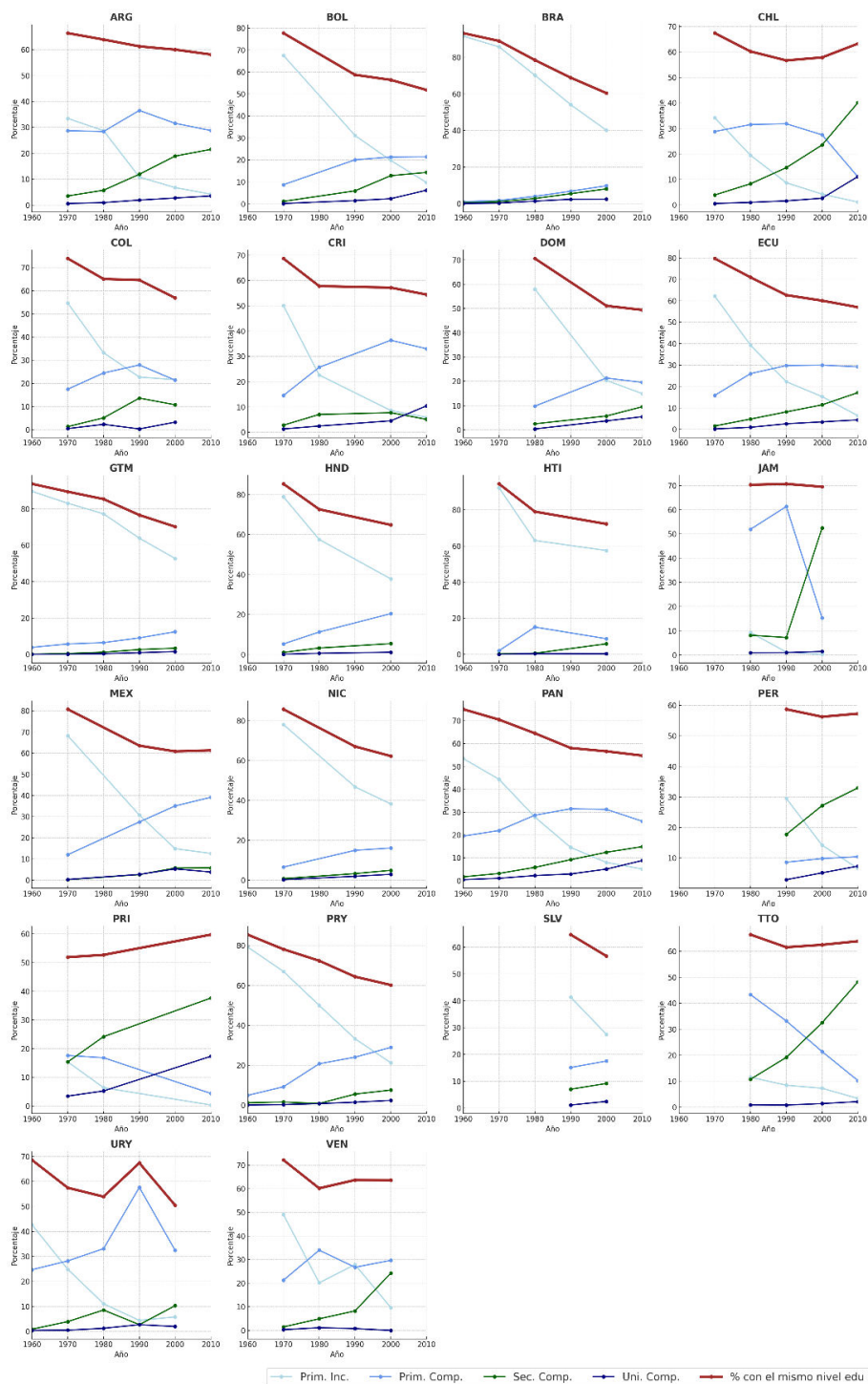
Panel B. Mujeres



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)

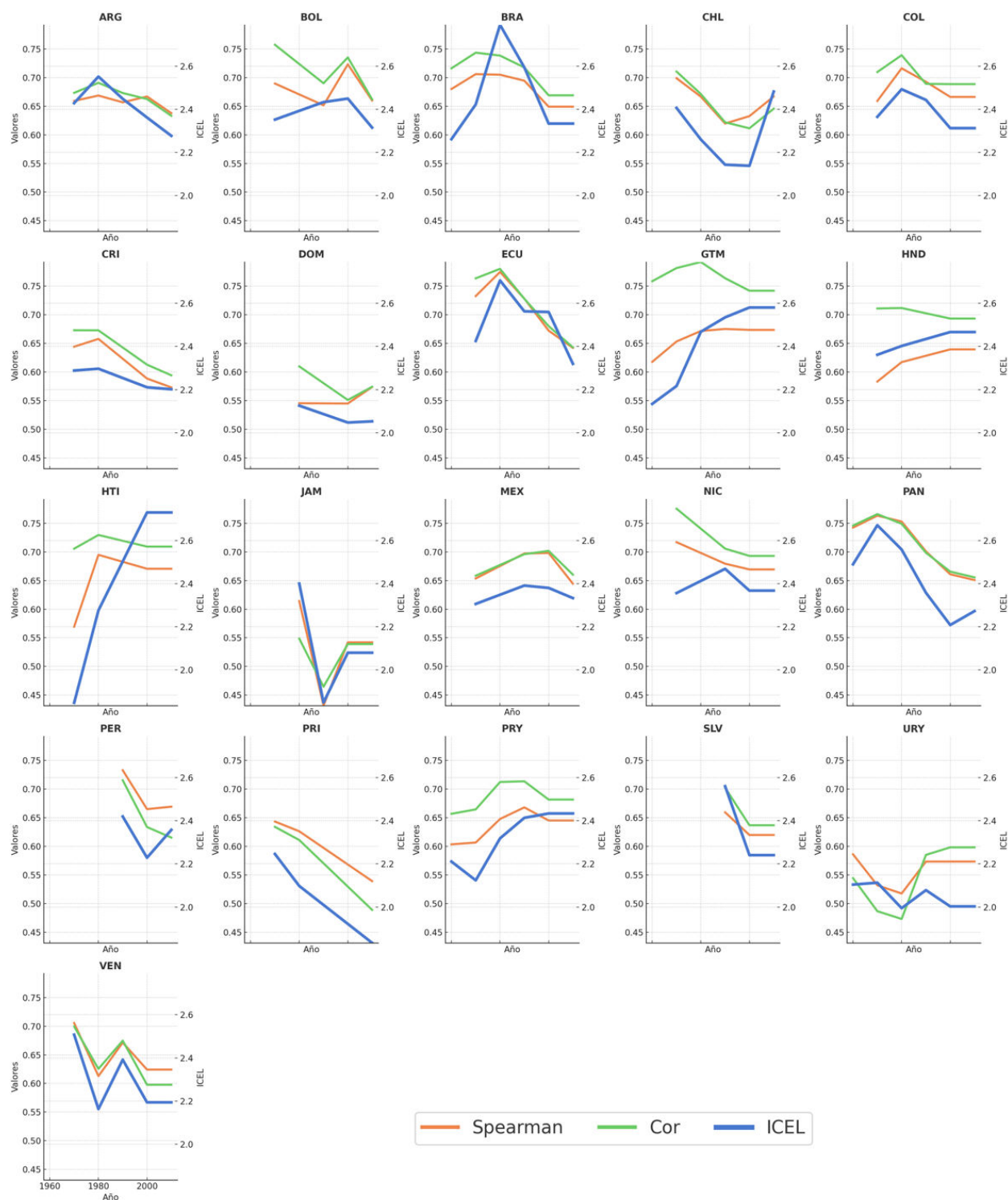
Figura A2. Porcentaje de parejas por nivel educativo – por país y por década de censo



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).

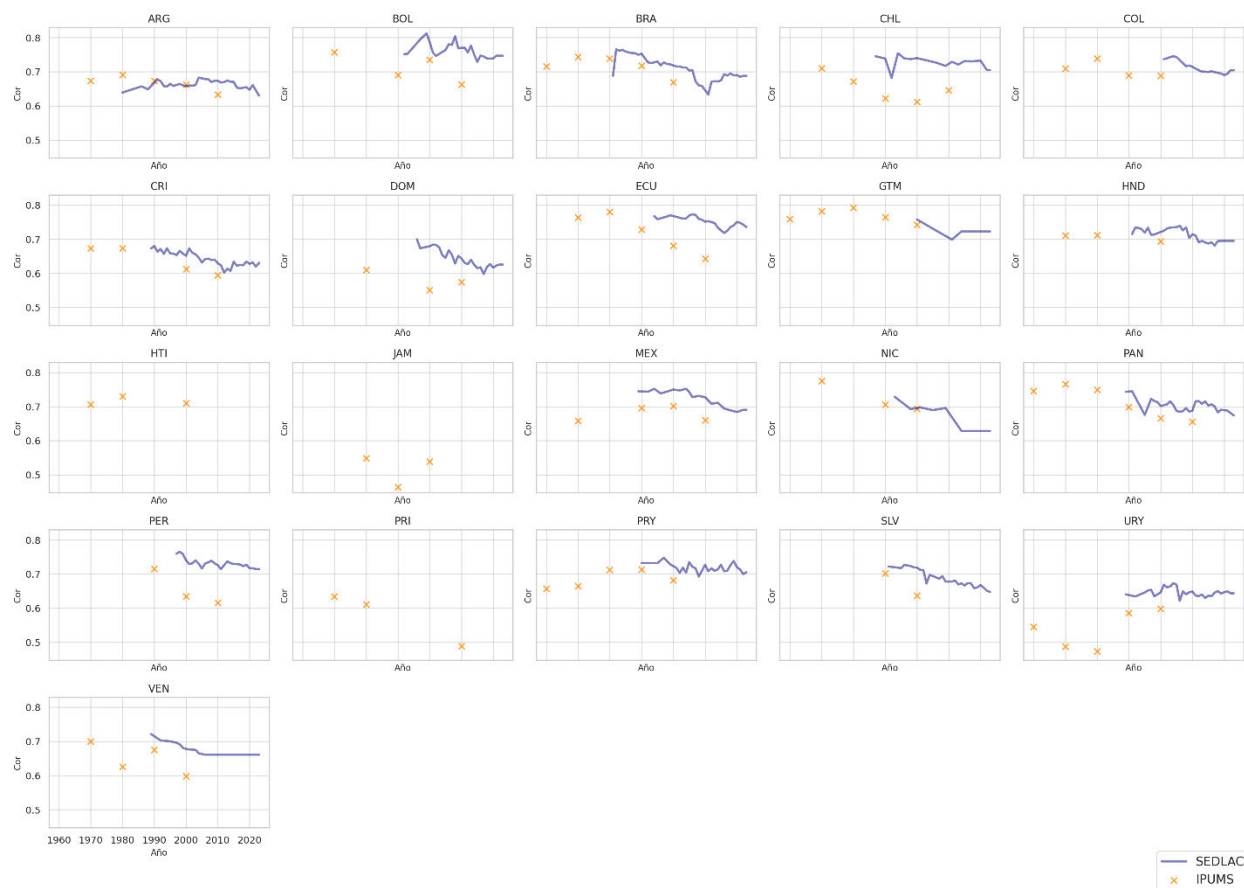
Figura A3. Evolución del ES según indicadores alternativos:



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).

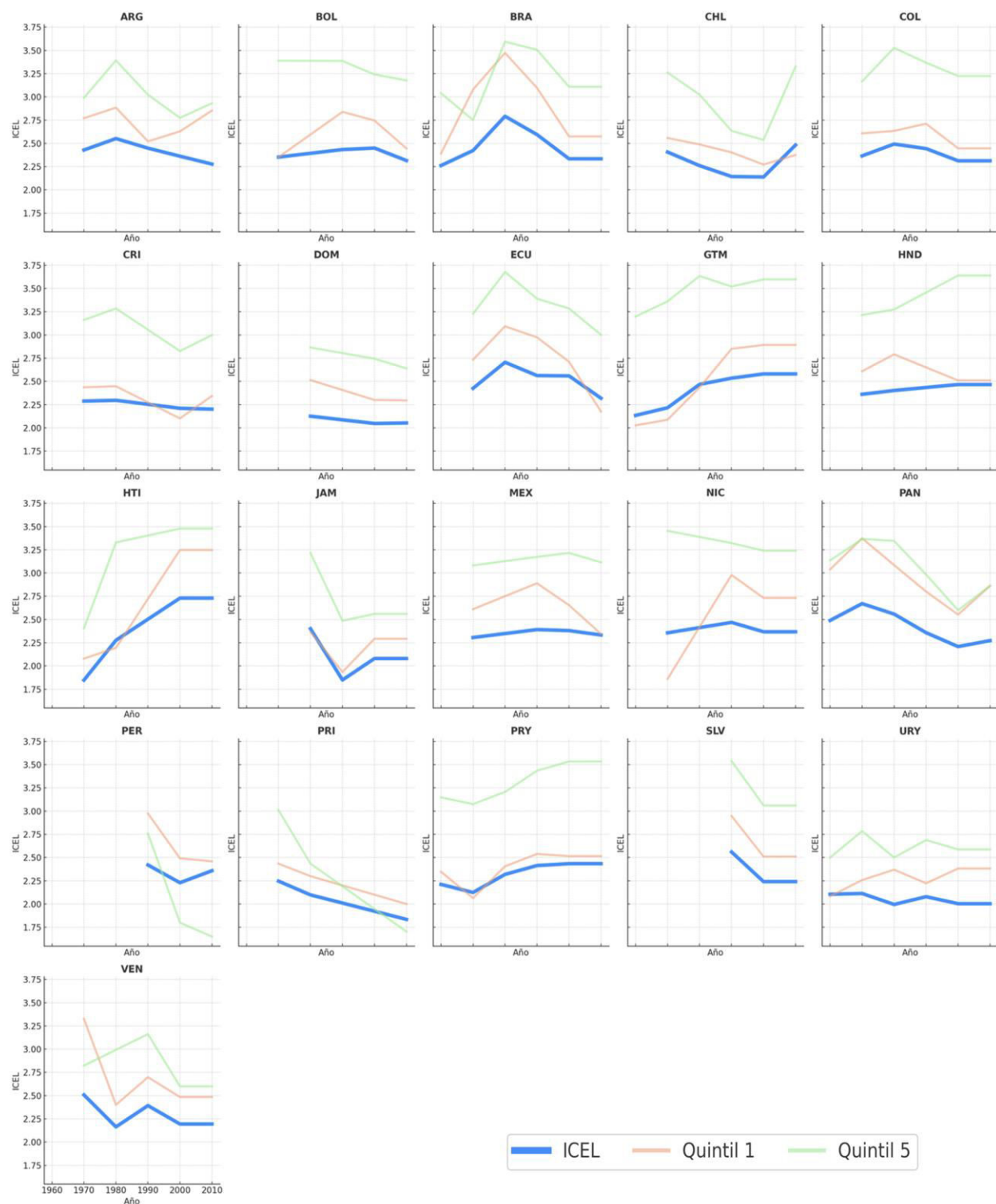
Figura A4. Coeficiente de correlación de Pearson según fuentes alternativas



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).

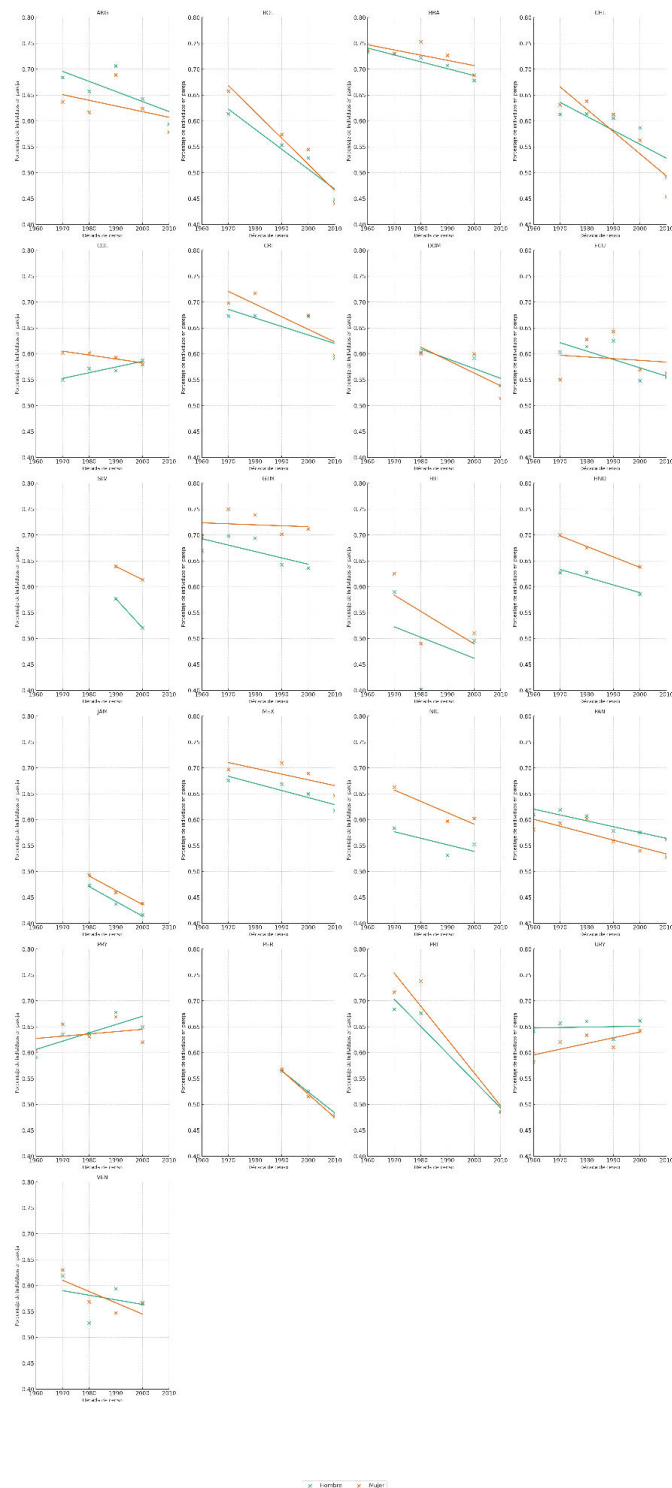
Figura A5. Evolución del ICEL



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).

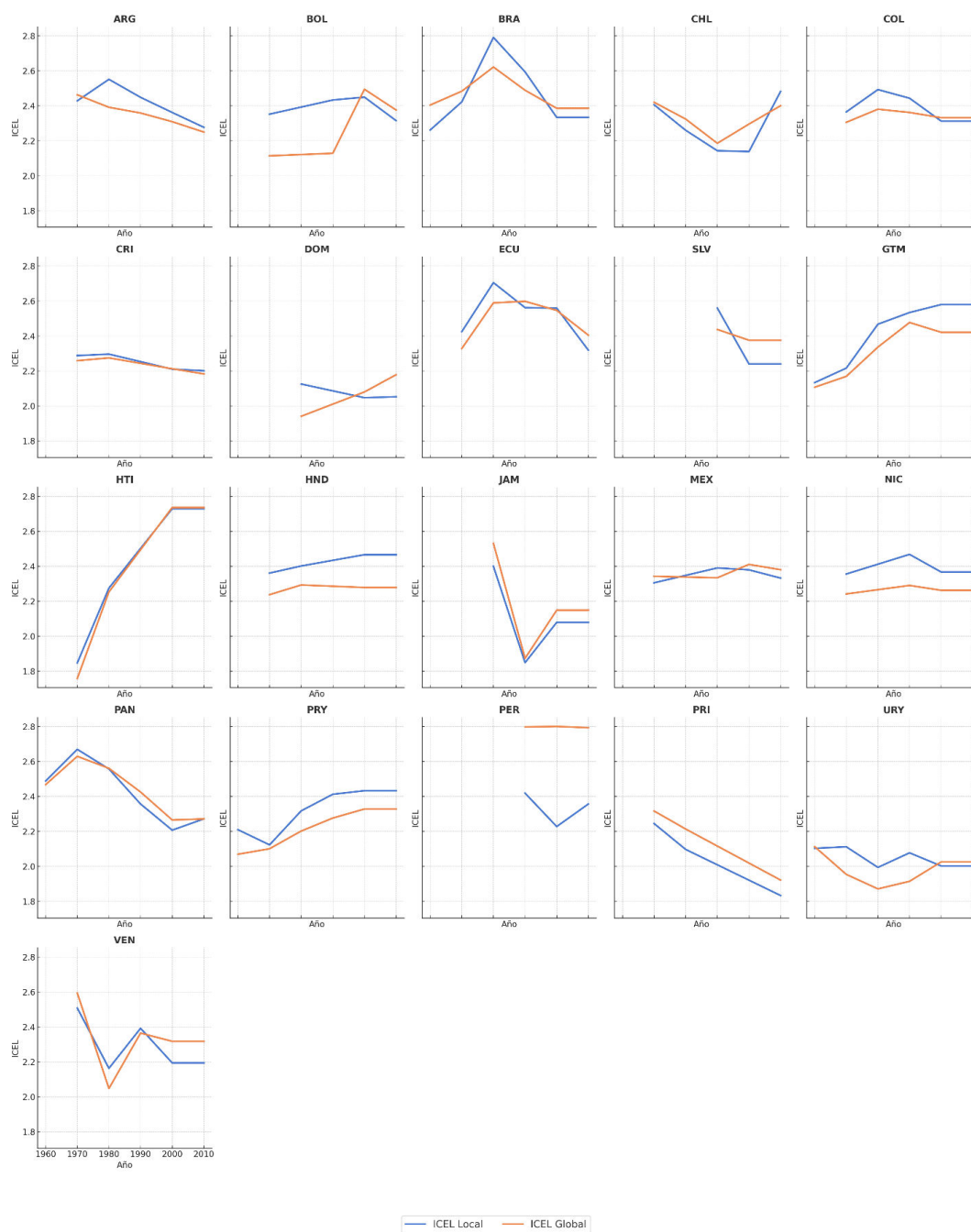
Figura A6. Porcentaje de individuos en pareja, por país y censo.



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)

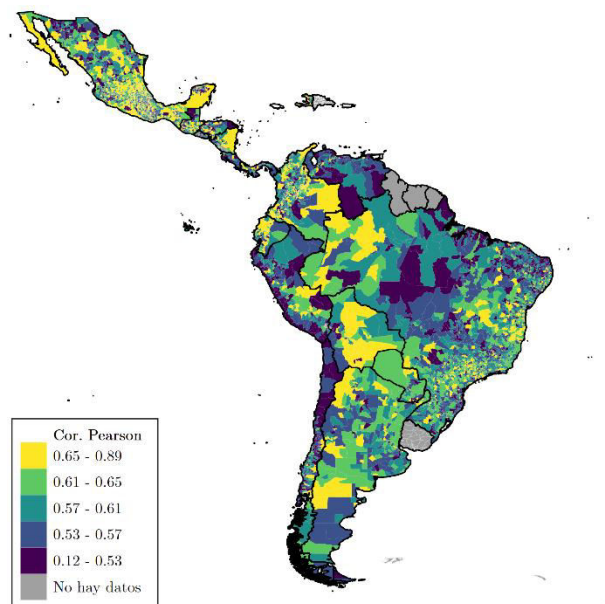
Figura A7. Evolución ICEG vs ICEL



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.

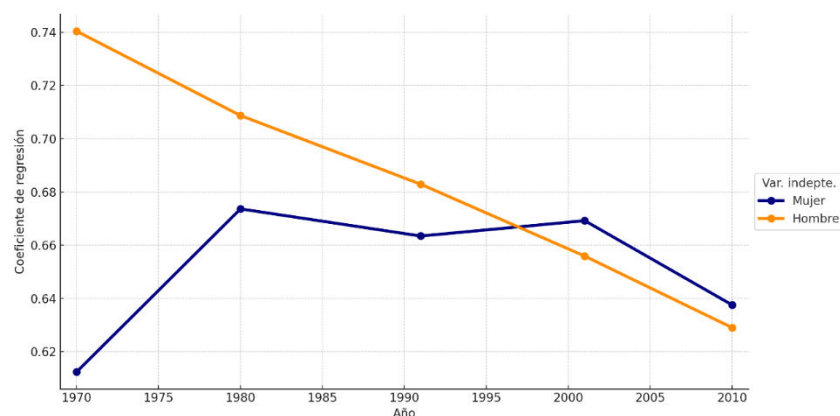
Apéndice II – Regresión y regresión inversa

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.

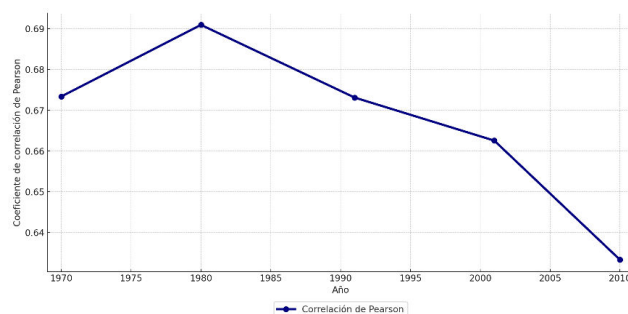
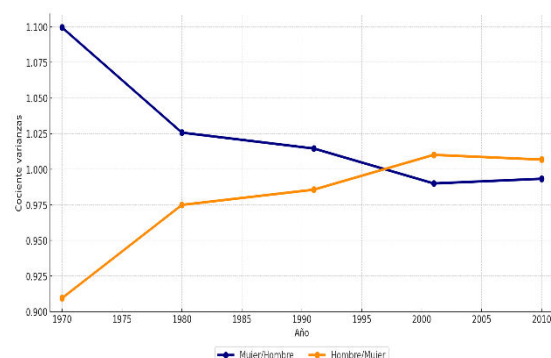
Fuente: elaboración propia en base a IPUMS (2020)

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

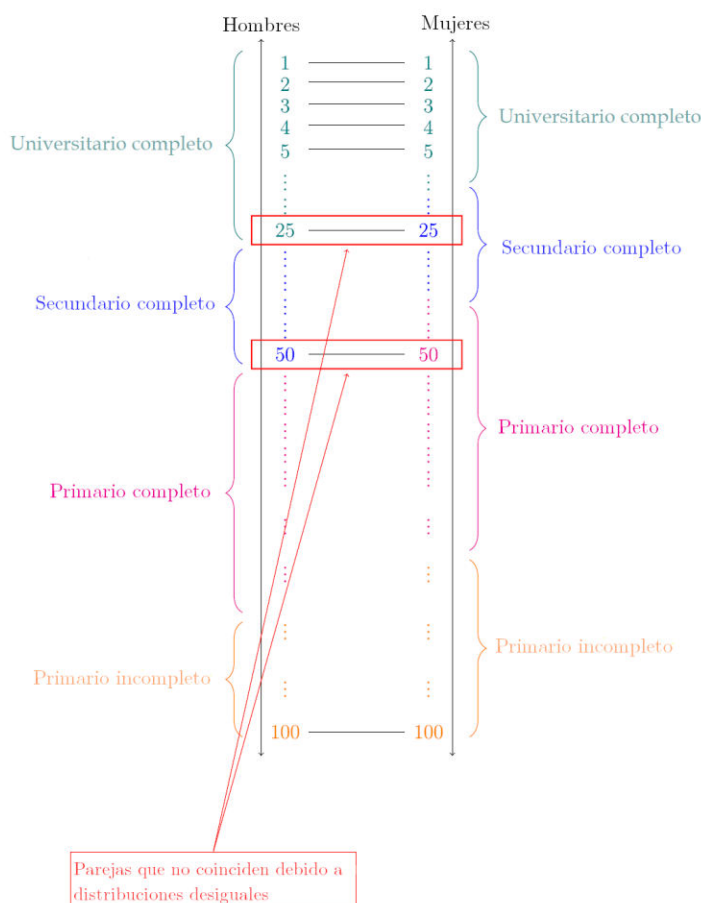
Una manera de evaluar la capacidad que tiene el indicador propuesto por Eika et al., (2019) para controlar completamente los cambios en las distribuciones educativas de hombres y mujeres es mediante la estimación del mismo en diferentes escenarios, en los cuales se asume un emparejamiento selectivo perfecto, 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 Argentina²⁹, 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

²⁹ 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.

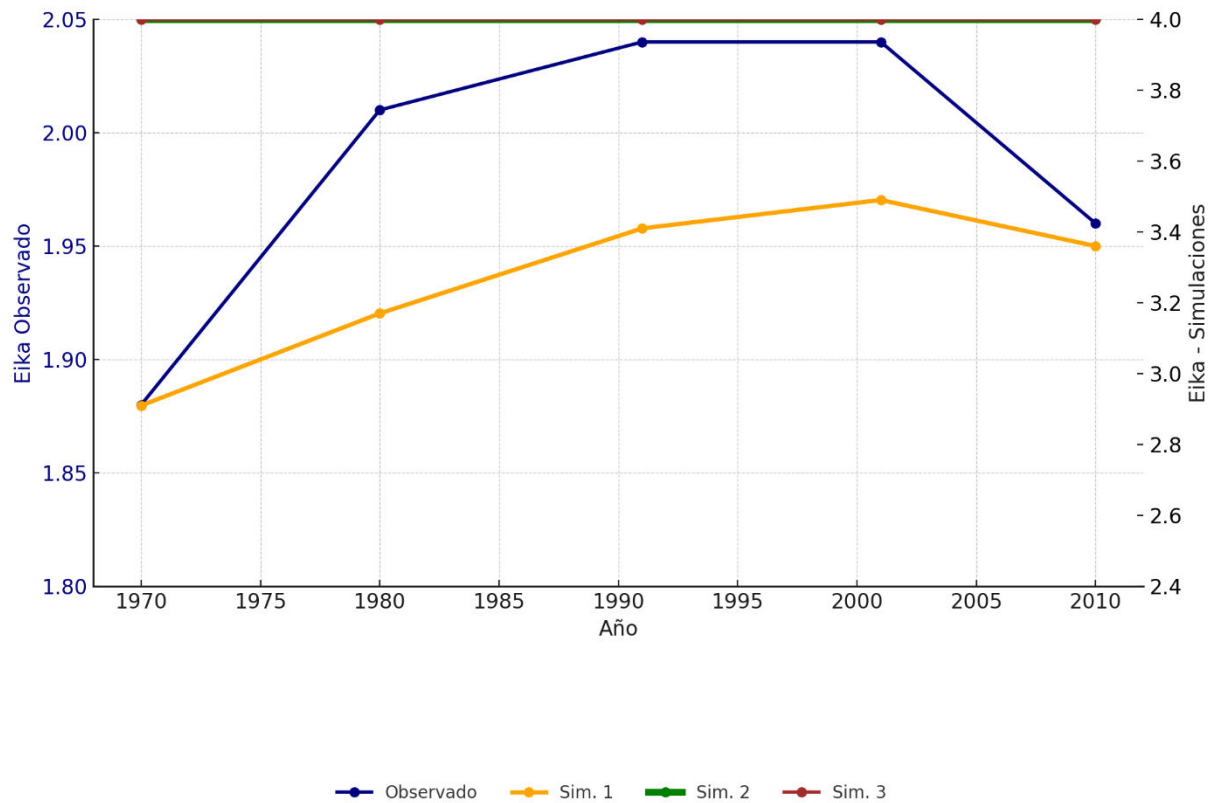
En la segunda simulación, se supondrá un emparejamiento selectivo perfecto y se permitirá que los tamaños de las categorías educativas varíen en el tiempo, pero se asumirá que no existen brechas de género en términos educativos.

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 máximo³⁰. 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

³⁰ 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)			
H\M	Alta	Baja	Total
Alta	0.50	0.20	0.70
Baja	0.00	0.30	0.30
Total	0.50	0.50	

M2. Matriz aleatoria		
H\M	Alta	Baja
Alta	0.35	0.35
Baja	0.15	0.15

M3. Matriz Eika		
H\M	Alta	Baja
Alta	1.43	0.57
Baja	0.00	2.00

Indicador Eika: 1.64

M1. Mat obs (Perfecta)			
H\M	Alta	Baja	Total
Alta	0.50	0.10	0.60
Baja	0.00	0.40	0.40
Total	0.50	0.50	

M2. Matriz aleatoria		
H\M	Alta	Baja
Alta	0.30	0.30
Baja	0.20	0.20

M3. Matriz Eika		
H\M	Alta	Baja
Alta	1.67	0.33
Baja	0.00	2.00

Indicador Eika: 1.81

M1. Mat obs (Perfecta)			
H\M	Alta	Baja	Total
Alta	0.50	0.00	0.50
Baja	0.00	0.50	0.50
Total	0.50	0.50	

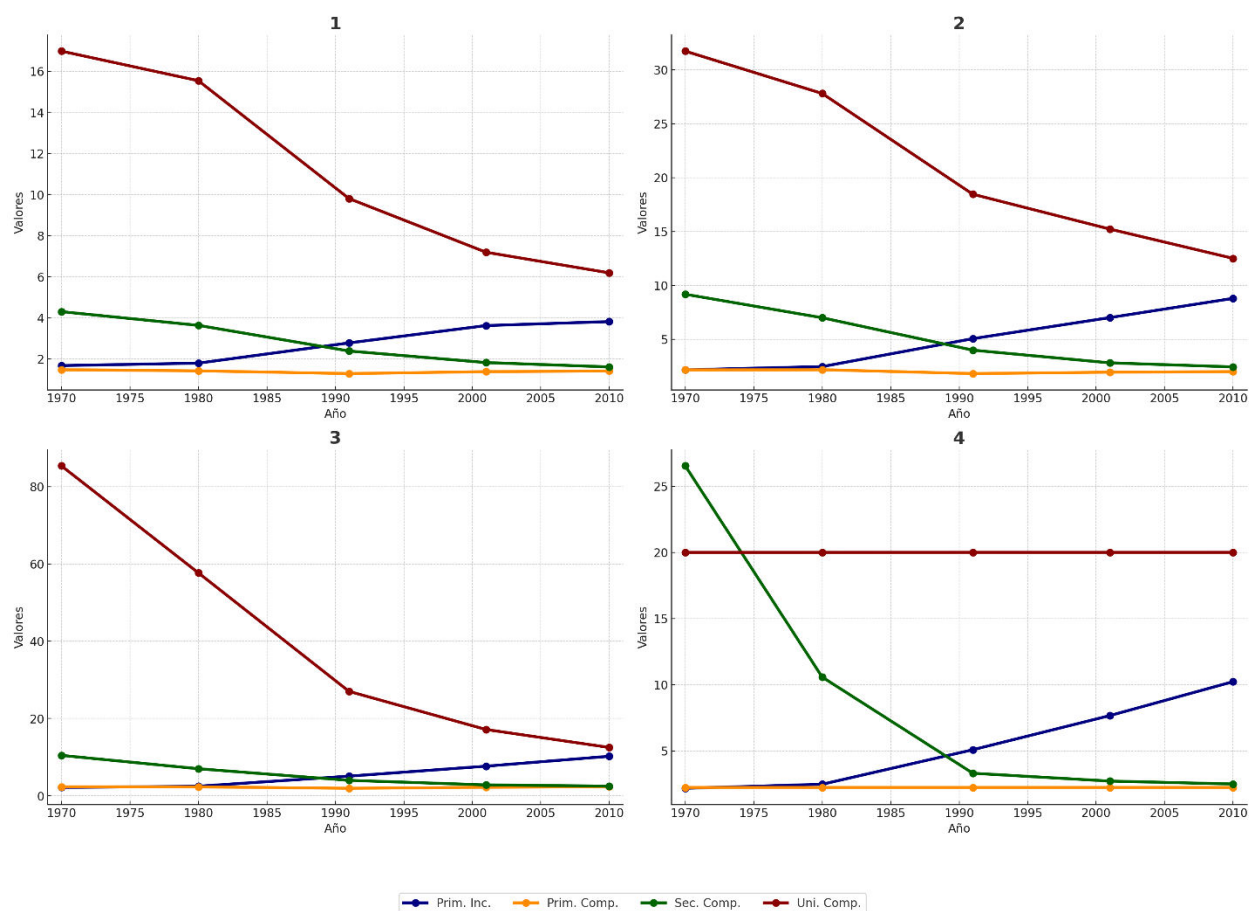
M2. Matriz aleatoria		
H\M	Alta	Baja
Alta	0.25	0.25
Baja	0.25	0.25

M3. Matriz Eika		
H\M	Alta	Baja
Alta	2	0
Baja	0	2

Indicador Eika: 2

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 límites³¹. 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 $p_{\text{aleatoria}}(e_o, e_a) = P(E_o = e_o) * P(E_a = e_a)$ ³². 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:

³¹Si 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.

³² 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.