

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.



ESTUDIO SOBRE LA MOVILIDAD INTERGENERACIONAL Y DIFERENCIAS
REGIONALES EN MÉXICO

TESINA

QUE PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

PRESENTA

RAFAEL MARTÍNEZ MARTÍNEZ

DIRECTOR DE LA TESINA: DR. RUBÉN IRVIN ROJAS VALDÉS

*Para mi jefa Lucy,
mi carnala Liz,
y para mi hija Sofía Abril.*

Agradecimientos

Al CIDE por haber cultivado en mí el deseo de superación personal y ser mejor cada día; así como también darme las facilidades para mi formación académica.

A mi asesor el Dr. Rúben Irvin Rojas Valdés por haberme brindado la confianza de trabajar con él, lo admiro y lo respeto por su gran conocimiento y calidad como persona.

Al maestro John Roberto Scott Andretta por los comentarios, correcciones y preguntas, que me hicieron reflexionar sobre el trabajo presentado.

A Lucy, Liz, Abril, Dani, Nise, y a toda mi familia por mantenerse unida en estos momentos tan complicados de salud por los que atravesamos.

Para Armando Fabian Lugo Peñaloza por ser un gran amigo incondicional y una persona con una gran calidad humana, su ayuda en aspectos personales y profesionales fue indispensable para la realización de este trabajo.

A Gilberto Castrejón Mendoza por ser parte de mi formación profesional, por ser un gran amigo y consejero en cuestiones personales y profesionales.

A Fede, Miguel, Thania, Dani, Antonio, Enrique, Lalo, Masa y Alonso, mis compañeros de la maestría en economía, por el apoyo y permitirme aprender de ellos.

También a todos los amigos y compañeros que de forma directa o indirecta hicieron posible que concluyera este trabajo.

Resumen

En los últimos años se ha estudiado la movilidad intergeneracional de México y de sus regiones. El principal problema que se encuentra para estimar la movilidad es la disponibilidad de datos. En este trabajo se estiman dos indicadores para medir la movilidad: la elasticidad intergeneracional del ingreso (EII); y la asociación intergeneracional del rango (AIR), tanto a nivel nacional y regional.

Para subsanar el problema de disponibilidad de datos, se emplea una encuesta de movilidad social en donde se tiene acceso a las condiciones socioeconómicas actuales de los entrevistados, además de información retrospectiva de su hogar de origen.

Se emplean técnicas de aprendizaje automático para seleccionar las características que resultan ser mejores predictores de los ingresos. Estas características permiten estimar los ingresos en ambas generaciones y establecer la relación de movilidad intergeneracional, donde el método de mínimos cuadrados en dos etapas en dos muestras fue la principal herramienta de estimación.

A diferencia de lo reportado en la literatura, se encuentra que existe movilidad nacional y regional, donde los individuos con menor (mayor) ingreso mejoran (empeoran) en cierta medida su situación económica respecto a su situación de origen. A nivel regional, el sur del país es el que experimenta las condiciones más desfavorables de movilidad, y la región norte es la que experimenta mejores condiciones de movilidad, lo cual coincide con la literatura enfocada en México.

Índice general

Índice de tablas

Índice de figuras

1. Introducción	1
2. Revisión de literatura	5
3. Modelos empíricos	7
3.1. Elasticidad intergeneracional del ingreso	8
3.2. Asociación intergeneracional del rango	13
3.3. Estimación Regional	14
3.4. Interpretación: EII, AIR	16
3.5. Muestreo bootstrap	17
4. Datos	19
4.1. Encuesta ESRU - EMOVI	19
4.1.1. Selección de variables para la estimación de ingresos	23
4.2. Encuestas ENIGH	25
5. Resultados y discusión	27
5.1. Elección de estimación	27
5.2. Movilidad nacional	29
5.3. Movilidad regional	33
5.4. Movilidad percibida	38
5.5. Alcance de resultados	40

6. Conclusiones	43
Bibliografía	45
Bases de datos	46
A. Estadísticas ESRU - EMOVI	47
B. Estadísticas ENIGH	50
C. Clasificaciones	59
D. Estimaciones	60
E. Relación de ingresos estimados	63

Índice de tablas

4.1. Intervalos de ingreso	20
4.2. Base de datos construida con los criterios <i>A</i> y <i>B</i>	22
4.3. Ingresos del hogar	26
5.1. Estimación de movilidad nacional	28
5.2. Estimación regional de movilidad	29
5.3. Probabilidad (%) de cambiar de quintil	32
5.4. Estimación regional de movilidad	33
5.5. Región norte: Probabilidad (%) de cambiar de quintil	35
5.6. Región centro-norte: Probabilidad (%) de cambiar de quintil	35
5.7. Región centro: Probabilidad (%) de cambiar de quintil	35
5.8. Región sur: Probabilidad (%) de cambiar de quintil	36
C.1. Codificación de ocupaciones, Clasificación Mexicana de Ocupaciones (CMO)	59
C.2. Codificación de años de educación	59
C.3. Codificación de regiones	59
D.1. Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango nacionales	60
D.2. Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango (criterio <i>A</i>) regionales. . .	61
D.3. Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango (criterio <i>B</i>) regionales. . .	62

Índice de figuras

3.1. Para cada individuo se construyen <i>gemelos sintéticos</i> que tengan las características del criterio <i>A</i> o el criterio <i>B</i> en la base del año t'	11
3.2. Regiones de México, de acuerdo a Banco de México (2011-2021)	15
4.1. Para cada individuo se construyen <i>gemelos sintéticos</i> que tengan las características del criterio <i>A</i> o el criterio <i>B</i> en la base del año t'	22
4.2. Precisión de las máquinas de soporte vectorial.	24
5.1. Relación entre el rango percentil promedio actual estimado y PP.	30
5.2. Densidad de ingreso nacional	31
5.3. Relación entre de rango percentil promedio actual y PPS.	34
5.4. Densidad del logaritmo del ingreso de la población por región, para E94-E96-E98 y E16-E18.	36
5.5. Modelo AIR estimado por regiones.	37
5.6. Percepción de movilidad de E17.	38
5.7. Percepción de movilidad muestra de E17 de acuerdo a intervalos de ingreso.	39
5.8. Modelo de AIR para: EUA, Dinamarca y Canadá.	40
5.9. El cuadro en rojo resalta la estimación de la EII, para los grupos de nacimiento 1981-85 y 1986-90.	42
A.1. Número de entrevistados seleccionados por estado y sexo, criterio A.	47
A.2. Número de entrevistados seleccionados por edad y sexo, criterio A.	47
A.3. Número de entrevistados seleccionados por intervalo de ingreso y sexo, criterio A.	48
A.4. Número de entrevistados seleccionados por estado y sexo, criterio B.	48
A.5. Número de entrevistados seleccionados por edad y sexo, criterio B.	49

A.6. Número de entrevistados seleccionados por intervalo de ingreso y sexo, criterio B.	49
B.1. Promedio de ingresos por edad para E94.	50
B.2. Promedio de ingresos por edad para E96.	50
B.3. Promedio de ingresos por edad para E98.	51
B.4. Histograma de ingreso E94 (logaritmo).	51
B.5. Histograma de ingreso E96 (logaritmo).	52
B.6. Histograma de ingreso E98 (logaritmo).	52
B.7. Densidad de ingreso empleando las muestras E94, E96 y E98.	53
B.8. Densidad de ingreso por región para E94.	53
B.9. Densidad de ingreso por región para E96.	54
B.10. Densidad de ingreso por región para E98.	54
B.11. Densidad de ingreso de la población para E94, E96 y E98.	55
B.12. Promedio de ingreso por edad y región para E94.	55
B.13. Promedio de ingreso por edad y región para E96.	56
B.14. Promedio de ingreso por edad y región para E98.	56
B.15. Densidad de ingreso de la población para E16 y E18.	57
B.16. Densidad de ingreso de la población para la unión de E16 y E18.	57
B.17. Densidad de ingreso de la población por región para la unión de E16 y E18.	58
B.18. Densidad de ingreso de la población E94-E96-E98, E16-E18.	58
E.1. Relación entre ingreso actual y PPS.	63
E.2. Relación entre el logaritmo de ingreso actual y PPS.	64
E.3. Relación de ingresos estimados.	65
E.4. Relación del logaritmo de ingresos estimados por región.	66

Introducción

La movilidad social puede entenderse como el cambio de un individuo en su situación socioeconómica. En ese sentido, se tiene que distinguir entre diferentes tipos de movilidad social. La movilidad horizontal, es aquella que se presenta al interior del estrato socioeconómico, mientras que la movilidad vertical o relativa, es aquella que se da al exterior del estrato socioeconómico.

En particular, la movilidad vertical intergeracional (movilidad intergeneracional) es aquella que sucede cuando un individuo logra una posición socioeconómica distinta a la que ocupaba su hogar de origen, la cual es el objetivo principal del presente trabajo.

Estudiar la movilidad en un país o región, es importante desde dos aristas. La primera en el sentido de justicia, pues la situación socioeconómica de una persona no debería de estar predeterminada por sus condiciones de origen. La segunda es la eficiencia, pues personas con potencial para incorporarse al mercado laboral podrían no llegar a los niveles requeridos de especialización, debido a sus condiciones de origen, ocasionando deficiencia en la asignación de oferta y demanda del mercado laboral, pues no necesariamente los más aptos están siendo asignados (Delajara y col., 2018).

Además, existe una estrecha relación entre desigualdad, movilidad y crecimiento económico (Delajara y col., 2018). Por lo anterior, es de importancia cuantificar en qué grado y en qué sentido se presenta movilidad en un país o región.

La movilidad intergeneracional puede ser medida de formas distintas, por ejemplo: ingreso familiar, ganancias individuales, clase social, estatus ocupacional y educación. En

este trabajo nos enfocamos en la dimensión del ingreso, por ser una dimensión directamente afectada por las políticas. Sin embargo, es evidente que todas estas dimensiones están interrelacionadas y el entendimiento de la movilidad social involucra estudiar dichas interrelaciones conjuntamente

La medición más común es la relación de ingresos entre los padres y sus hijos mayores. En este trabajo se utiliza el término *proveedor principal (PP)* para sustituir el término de padres, y el de *individuo actual* para sustituir el de hijos. Lo anterior para incluir tanto a hombres como mujeres en ambas situaciones.

El estándar para medir la movilidad es establecer una regresión lineal del logaritmo del ingreso (y^a) de los individuos actuales, sobre el logaritmo del ingreso (y^p) de los PP. Al parámetro de relación se le denomina elasticidad intergeneracional del ingreso (EII).

La EII no mide la relación causal entre los ingresos a través de las generaciones, es la medición de la persistencia intergeneracional,¹ que provee una medida de las desigualdades intergeneracionales (Blanden, 2019).

Una desventaja de las EII es que esta se ve afectada por cambios en la distribución del ingreso. Así, si la desigualdad está creciendo, de modo que sea mayor en la generación del individuo actual que en la del PP, este cambio en la desigualdad se manifestará automáticamente como una mayor EII.

Una medida alternativa a la EII, es la asociación intergeneracional del rango (AIR).² Esta medida captura la correlación en los rangos percentiles de ingresos de forma completamente independiente de cualquier cambio en la distribución de ingresos entre las dos generaciones.³

¹ Si el valor es cero describe una situación donde el ingreso nacional es igualmente distribuido en toda la población, si es uno, describe que todo el ingreso nacional es retenido por un solo individuo (Blanden, 2019).

² Algunos autores llaman a esta estimación rango-rango. En este trabajo se utiliza AIR, siguiendo la referencia que hace Chetty y col. (2014).

³ Una AIR de 1 indica que los individuos actuales imitan exactamente el rango percentil de sus PP en la distribución del ingreso.

La estimación de la movilidad intergeneracional utilizando EII o AIR, depende de la calidad de los datos disponibles. Idealmente, se podrían estimar dichos indicadores si se dispusiera de datos en panel largo de dos o más generaciones. Así, sería posible analizar la relación entre los ingresos observados de la generación actual con los ingresos observados de la generación de origen. Sin embargo, hay poca disponibilidad y baja calidad en los datos, sobre todo en países en desarrollo.

En el presente trabajo se implementan las estimaciones de la EII, y de la AIR. Se utiliza la encuesta ESRU-EMOVI-2017 (E17) para obtener información de los individuos actuales y de los PP. Con el fin de aumentar los datos disponibles, se utilizan tanto datos de hombres como mujeres.

Para solventar la falta de datos de individuos actuales, se realiza una asociación de información, con ayuda de algoritmos de aprendizaje automático, entre la E17 y las encuestas ENIGH 2016 (E16) y 2018 (E18). Para estimar los ingresos de la generación de origen se utilizan datos de las encuestas ENIGH 1994 (E94), 1996 (E96), 1998 (E98), y se aplica el método de mínimos cuadrados en dos etapas en dos muestras para obtener las estimaciones de movilidad.

Los resultados apuntan a que existe movilidad en México y sus regiones, pero esta movilidad tiene unas características específicas, los individuos con menor (mayor) ingreso mejoran (empeoran) en cierta medida su situación económica respecto a su situación de origen. A nivel regional, el sur del país es el que experimenta las condiciones más desfavorables de movilidad, y la región norte es la que experimenta mejores condiciones de movilidad, lo cual coincide a nivel cualitativo con la literatura enfocada en México. Se ha creado un apéndice en línea, que contiene las implementaciones que se elaboraron con ayuda de R y Python.⁴

⁴ La dirección es la siguiente: <https://rafneta.github.io/TesinaCIDEMERMM/index.html>

La organización del documento es la siguiente: la revisión de los estudios de movilidad para México, que se enfocan en la EII y AIR, se encuentran en el capítulo 2. Las estrategias de estimación y el planteamiento del problema se presentan en el capítulo 3, un resumen de las características de los datos utilizados y la técnica de construcción de los mismos, se encuentra en el capítulo 4, el capítulo 5 contiene los resultados y la discusión, finalmente en el capítulo 6 se presentan las conclusiones.

Revisión de literatura

A continuación se presentan algunos de los resultados que se han obtenido para México.

En el estudio de Delajara y Graña (2018), los autores toman datos de la encuesta sobre movilidad ESRU - EMOVI - 2011, emplean regresiones rango-rango (AIR) para estimar movilidad regional. En este trabajo se concluyen que el grado de movilidad en las regiones Norte y Centro-Norte del país es mayor que el promedio nacional, mientras que en la región Sur es menor. En particular su implementación hace uso de índices de bienestar (ocupación, educación, salud) en lugar de tomar de los ingresos.

En Campos-Vázquez y col. (2020) se emplean datos de ingresos obtenidos de censos entre 1990 y 2015, haciendo una relación *sintética* entre padres e hijos supuestos, pues no se tiene la información real. Para construir dicha relación los autores usan la información sobre fechas y el estado de nacimiento de los hijos, así como los ingresos de los padres, asociando los padres por su cohort de nacimiento. Su conclusión es que EII es de 0.5 para México, hablaremos más sobre sus resultados en el capítulo 5. En cuanto al análisis regional sus resultados son consistentes con Delajara y Graña (2018).

En Monroy-Gómez-Franco y Corak (2020) se hace énfasis que en México y otros países con economías en desarrollo, hay poca o nula información existente sobre datos de ingresos de padres e hijos. Así, crean índices a partir de características a niveles de hogar y vecindad, sus resultados también son consistentes con Campos-Vázquez y col. (2020), reportan una disminución del 12 % en términos de movilidad social hacia arriba en el Sur (respecto al promedio nacional), contra un incremento del 10 % para la Ciudad de México.

En un estudio publicado recientemente, Báez (2021) reporta la implementación del método de mínimos cuadrados en dos etapas en dos muestras (método implementado en este trabajo) para estimar la EII, así como regresión rango-rango, a nivel nacional y regional. En este trabajo se reporta una menor movilidad que la de algunos trabajos previos. En el aspecto regional, su información es consistente con la de los trabajos antes mencionados, se hablará más de estos resultados en el capítulo 5.

Modelos empíricos

Para analizar la relación intergeneracional de los ingresos en México y en sus regiones se emplean los siguientes dos modelos econométricos.

La estructura general del modelo para la EII es:

$$\log(y^a) = \beta_0 + \log(y^p)\beta + X\sigma + v$$

mientras que para AIR se tiene:

$$R(y^a) = \alpha + R(y^p)\rho + w$$

Donde y^a hace referencia a los ingresos del individuo actual, y y^p hace referencia a los ingresos del PP del individuo actual. El valor estimado de los parámetros β , ρ , indica el grado de movilidad económica. Finalmente, la función R es el rango percentil.

El principal problema para la estimación de los parámetros de estos modelos es que no se cuentan con los ingresos del PP, pues las bases de datos existentes en México no hacen un seguimiento de los individuos en el tiempo. Por tal motivo, se tiene que construir una estimación del ingreso y^p .

Para ello se toma la encuesta ESRU - EMOVI - 2017 (E17), que cuenta con información de los entrevistados y de las características (ocupación, educación, sexo, residencia, etc.) de su PP. Tomando esta información, se realiza una búsqueda de individuos en encuestas previas que tengan características semejantes a las del PP, y así, realizar una estimación del ingreso y^p , en particular, se toman las encuestas ENIGH 1994 (E94), 1996 (E96) y 1998

(E98) para dicho propósito.

Si bien la encuesta E17 proporciona información importante para la estimación de los ingresos y^p , tiene una desventaja, los ingresos y^a no están indicados de forma puntual, sino que se registran por intervalos de ingreso familiar. Por lo cual se tiene que aplicar una estrategia de estimación puntual para y^a para lograr esto, se toma información adicional de las encuestas ENIGH 2016 (E16) y ENIGH 2018 (E18).

Una vez resueltas las estimaciones de ingresos, se pueden estimar los parámetros de los modelos, y así, obtener información sobre la movilidad económica intergeneracional en México y sus regiones.

En las siguientes secciones se indican con detalle las especificaciones de cada modelo.

3.1. Elasticidad intergeneracional del ingreso

La EII, mide la persistencia entre los ingresos de un individuo y su descendencia. El modelo es

$$\log(y^a) = \beta_0 + \log(y^p)\beta + X\sigma + v \quad (3.1)$$

Donde y^a hace referencia a los ingresos del individuo actual, y y^p hace referencia a los ingresos de PP del individuo actual, X son las variables de control. La estimación $\hat{\beta}$ es una medida de la persistencia en el estatus económico, si $\hat{\beta} \approx 0$ existe alta movilidad, $\hat{\beta} \approx 1$ indica una muy baja movilidad.

Como se mencionó en la introducción del capítulo, el problema para la estimación de los parámetros del modelo, es que no se cuenta con los ingresos del PP, las bases de datos existentes en México no hacen un seguimiento de los individuos en el tiempo. Por tal motivo, se tiene que construir una estimación del ingreso y^p .

Una solución a dicho problema lo brinda el método de Mínimos Cuadrados en dos

Etapas en dos Muestras (MC2E2M). Este método propone estimar, en una primera etapa, los ingresos del PP haciendo uso de una muestra de datos que contenga información semejante a las características de los PP, y posteriormente utilizar dichas estimaciones en una segunda etapa.

En la segunda etapa se realiza la estimación de los parámetros del modelo (3.1), donde se toma la información de los y^a de una segunda muestra de datos, y se hace uso de las estimaciones \hat{y}^p de la primer etapa. Se sigue en parte a Mocetti (2007), para implementar el modelo MC2E2M.⁵

$$\begin{aligned}
 \text{(Etapa 1)} \quad & \log(y^p) = Z\delta + S^p\sigma^p + A^p(e^p) + u, & (3.2) \\
 \text{(Etapa 2)} \quad & \log(y^a) = \widehat{\log(y^p)}\beta + S^a\sigma^a + A^a(e^a) + v, \\
 & \widehat{\log(y^p)} = Z\hat{\delta} + S^p\hat{\sigma}^p + \hat{A}^p(e^p)
 \end{aligned}$$

donde A^p es un polinomio de segundo grado que depende de la edad e^p del PP, A^a es un polinomio de segundo grado que depende de la edad e^a del individuo actual, S^a es el control, en nuestro caso el sexo del individuo actual, mientras que S^p es el sexo del PP, u y v son los errores, y Z son las características de los PP, en este caso: educación y ocupación.

Para obtener la información de los PP (variables Z , S^p , e^p), se utiliza la encuesta E17, que cuenta con información de los entrevistados (muestra de datos de la segunda etapa) y de las características (ocupación, educación, sexo, edad) de su PP, en particular dichas características se recolectan preguntando por la situación del entrevistado a los 14 años de edad.

Tomando esta información, se busca a individuos en encuestas previas (proveedor principal sintético PPS) que tengan características semejantes a las del PP para ejecutar la etapa 1 del MC2E2M. Los entrevistados seleccionados⁶ de E17 tienen una edad promedio de 36 años, entonces en promedio en 1995 tenían 14 años. Por tal motivo se seleccionan encuestas lo más cercanas a 1995. Se toman las encuestas E94, E96 y E98 para dicho

⁵ Se presenta una especificación general del modelo. Más adelante se indica de forma específica.

⁶ La descripción precisa de los datos se encuentra en el capítulo 4.

propósito (muestra de datos de la primer etapa).⁷

Al aplicar el método de MC2E2M puede que la estimación presente desviaciones debidas a diferentes factores, siguiendo a Báez (2021) y Blanden (2019), las posibles desviaciones son:

- **Desviación por atenuación (DA).** Puede que la estimación presente desviación hacia abajo por utilizar solo una sección cruzada para la estimación de los ingresos de los PP, lo cual producirá una inconsistencia de los estimadores. Para reducir este error se construye un promedio de estimación de ingresos de los PP con las datos de las encuestas E94, E96 y E98
- **Desviación por ciclo de vida (DC).** Esta es ocasionada por la observación del ingreso en una edad determinada. Debido a que la muestra de datos con la que se trabaja (subconjunto de E17) es pequeña ($N = 1422$), no será posible establecer una comparación tomando individuos con diferentes grupo de edades, entonces se tiene que tener en cuenta que los resultados están afectados por la DC
- **Desviación por selección de muestra (DS).** Puede surgir desviaciones debidas a la forma en cómo se recolectan los datos (información auto reportada, información dada por terceros, etc.), esto impacta en las hipótesis para la consistencia de los estimadores, para reducir esto se realiza muestreo con remplazo de los datos disponibles, además del promedio de estimación de ingresos de los PP con las datos de las encuestas E94, E96, y E98.

La encuesta E17 proporciona información fundamental para la estimación de los ingresos y^p , pero tiene una desventaja, los ingresos y^a no están indicados de forma puntual, sino que se registran por intervalos de ingreso familiar. Los ingresos están distribuidos en 7 intervalos, lo cual no hace posible la aplicación directa de la segunda etapa de MC2E2M.

Con lo cual surge la necesidad de aplicar una estrategia de estimación puntual para y^a . Para esto, se tomaron algunas de las características (sexo, edad, educación, ocupación y región de residencia) del entrevistado seleccionado de la encuesta E17, y se buscaron individuos en bases de datos adicionales (que denominamos *gemelos sintéticos*) que

⁷ Se toman estas tres encuestas por un detalle técnico que más adelante se aclara.

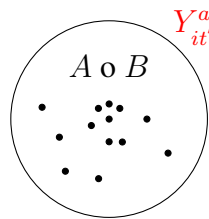
compartieran dichas características con el entrevistado.

Estos gemelos sintéticos tienen ingreso puntual registrado en su base de datos correspondiente, entonces relacionando a cada entrevistado seleccionado de E17 con su conjunto de gemelos sintéticos, se puede estimar el ingreso puntual para cada entrevistado de E17.

Con un primer criterio de selección (Criterio *A*), además de las características previamente mencionadas, se permite que los gemelos sintéticos no tengan exactamente la misma edad a la del entrevistado, adicionalmente no se pide ninguna restricción al ingreso del gemelo. Con un segundo criterio (Criterio *B*) se impone la restricción de que el gemelo sintético tenga un ingreso que pertenezca al mismo intervalo reportado por el entrevistado.

Para tener una mayor cantidad de entrevistados con gemelos sintéticos asignados, se buscaron dichos gemelos sintéticos en las encuestas E16 y E18. Es decir, para un entrevistado i , con gemelos sintéticos encontrados en la encuesta del año $t' \in \{2016, 2018\}$, que reside en la región⁸ $r_k \in \{r_1, \dots, r_4\}$, se encontró un conjunto $Y_{it'}^a$ utilizando el criterio *A* y otro conjunto utilizando el criterio *B*. La Figura 3.1 muestra un esquema de la idea de esta construcción.⁹

Figura 3.1. Para cada individuo se construyen gemelos sintéticos que tengan las características del criterio *A* o el criterio *B* en la base del año t'



Fuente: Elaboración propia

Ahora se puede establecer de manera precisa el modelo de estimación para la EII nacional. Como se tienen tres bases de datos $t_j \in \{t_1 = 1994, t_2 = 1996, t_3 = 1998\}$ para estimar los ingresos del PP, se ha nombrado *modelo EII nacional fijo* a la siguiente especificación:

⁸ r_1 región norte, r_2 región centro-norte, r_3 región centro, r_4 región sur.

⁹ La información específica sobre la construcción de los conjuntos de gemelos sintéticos se encuentra en el capítulo 4.

Modelo EII nacional fijo (ME_{nf})

$$\begin{aligned}
 \text{(Etapa 1)} \quad & \log(y_{it_j}^p) = c_{0t_j} + Z_{lt_j} \delta_{t_j} + S_{it_j}^p \sigma_{t_j}^p + A_{t_j}^p(e_{it_j}^p) + u_{lt_j} \\
 \text{(Etapa 2)} \quad & \log(\hat{y}_i^a) = \beta_{0t_j} + \widehat{\log(y_{it_j}^p)} \beta_{t_j} + S_i^a \sigma^a + A_{t_j}^a(e_i^a) + v_{it_j} \\
 & Z_{lt_j} \delta_{t_j} = \sum_m E_{lm} \delta_{t_j E_m} + \sum_o O_{lo} \delta_{t_j O_o} \\
 & \widehat{\log(y_{it_j}^p)} = \hat{c}_{0t_j} + Z_{it_j} \hat{\delta}_{t_j} + S_i^p \hat{\sigma}_{t_j}^p + \hat{A}_{t_j}^p(e_i^p) \\
 & \hat{y}_{it_j}^a \in \left\{ \bar{Y}_{it_1}^a, \bar{Y}_{it_2}^a \right\}_C
 \end{aligned}$$

donde l denota al individuo de la encuesta t_j , E_m es la categoría de educación, O_o es la categoría de ocupación,¹⁰ $\bar{Y}_{it'}^a$ hace referencia al promedio aritmético del conjunto de ingresos de los gemelos sintéticos, $C \in \{A, B\}$ indica si los gemelos sintéticos se construyen con el criterio A o con el criterio B . Se agrego el control de sexo S de individuo. Cabe recordar que el parámetro de interés es β_{t_j} .

Con esta especificación se realizaron seis estimaciones de β , una para cada año t_j y por cada criterio de construcción de gemelos sintéticos C . Las estimaciones se encuentran en la Tabla D.1 del apéndice D.

Como se ha comentado existe una DA, por ello se recurre a realizar una nueva especificación del método MC2E2M, donde la etapa 1 se sustituye por el promedio de primeras etapas del modelo (ME_{nf}) para los años 1994, 1996, 1998. Tomando esto en cuenta, la especificación para el *modelo de EII nacional promedio* es:

¹⁰ Los detalles de estas variables se encuentran el capítulo 4.

Modelo EII nacional promedio (ME_{np})

$$\begin{aligned}
 \text{(Etapa 2)} \quad \log(\hat{y}_i^a) &= \beta_0 + \widehat{\log(y_i^p)}\beta + S_i^a\sigma^a + A^a(e_i^a) + v_i \\
 \widehat{\log(y_i^p)} &= \frac{1}{3} \left(\widehat{\log(y_{it_1}^p)} + \widehat{\log(y_{it_2}^p)} + \widehat{\log(y_{it_3}^p)} \right) \\
 \hat{y}_i^a &\in \left\{ \bar{Y}_{it_1}^a, \bar{Y}_{it_2}^a \right\}_C
 \end{aligned}$$

Con esta especificación se realizaron dos estimaciones de β , una por cada criterio de construcción de gemelos sintéticos C . Las estimaciones se encuentran en el Tabla D.1 del apéndice D y la discusión de resultados en el capítulo 5.

3.2. Asociación intergeneracional del rango

La AIR tiene como especificación

$$R(y^a) = \alpha + R(y^p)\rho + w \quad (3.3)$$

al igual que antes, y^a hace referencia a los ingresos de la individuo actual, y y^p hace referencia a los ingresos del PP del individuo actual, w es el error. El operador R hace referencia al cálculo del rango percentil.¹¹

La estimación $\hat{\rho}$ es una medida de la persistencia en el estatus económico, si $\hat{\rho} \approx 0$ existe alta movilidad, $\hat{\rho} \approx 1$ indica una muy baja movilidad. La diferencia entre la interpretación de ρ y β se establece al final de este capítulo 5.

Para estimar los ingresos actuales puntuales y^a se procede como en la estimación de EII. Para estimar y^p se utiliza la primera etapa del método de MC2E2M y una transformación exponencial. Se realiza tanto la estimación nacional fija como la promedio, en el mismo sentido que en la estimación de la EII. Con estas consideraciones, se presentan las especificaciones del modelo AIR.

¹¹ Dado un conjunto de datos ordenado $D = \{d_i\}$, el rango percentil de x con base en el conjunto D , es el porcentaje de datos del conjunto D que son menores o iguales a x .

El modelo de AIR nacional fijo es:

Modelo AIR nacional fijo (MR_{nf})

$$\begin{aligned} R(\hat{y}_i^a) &= \alpha_{t_j} + R(\hat{y}_{it_j}^p)\rho_{t_j} + w_{t_j} \\ \hat{y}_i^a &\in \left\{ \bar{Y}_{it_1}^a, \bar{Y}_{it_2}^a \right\}_C \\ \hat{y}_{it_j}^p &= \exp\left(\widehat{\log(y_{it_j}^p)}\right) \end{aligned}$$

El modelo de AIR nacional promedio es:

Modelo AIR nacional promedio (MR_{np})

$$\begin{aligned} R(\hat{y}_i^a) &= \alpha + R(\hat{y}_i^p)\rho + w \\ \hat{y}_i^a &\in \left\{ \bar{Y}_{it_1}^a, \bar{Y}_{it_2}^a \right\}_C \\ \hat{y}_i^p &= \frac{1}{3} (\hat{y}_{it_1}^p + \hat{y}_{it_2}^p + \hat{y}_{it_3}^p) \\ \hat{y}_{it_j}^p &= \exp\left(\widehat{\log(y_{it_j}^p)}\right) \end{aligned}$$

Con estas especificaciones se realizaron las estimaciones de ρ^{12} y α . Las estimaciones se encuentran en el Tabla D.1 del apéndice D y la discusión de resultados en el capítulo 5.

3.3. Estimación Regional

Desde el primer trimestre del 2011, el Banco de México presenta los indicadores y el análisis sobre la evolución trimestral de la actividad económica, el comportamiento de los precios y las expectativas de los agentes económicos de las regiones (Banco de México, 2011-2021):

- **Norte.** Baja California, Chihuahua, Coahuila, Nuevo León, Sonora y Tamaulipas.

¹² El rango percentil se obtiene con base en el total de ingresos de la encuesta correspondiente, en el caso de los ingresos actuales se calculó en la unión de E16 y E18.

- **Centro-Norte.** Aguascalientes, Baja California Sur, Colima, Durango, Jalisco, Michoacán, Nayarit, San Luis Potosí, Sinaloa y Zacatecas.
- **Centro.** Ciudad de México, Estado de México, Guanajuato, Hidalgo, Morelos, Puebla, Querétaro y Tlaxcala.
- **Sur.** Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

Debido a que se tiene esta clasificación de regiones, se decide establecer la misma división para el análisis de movilidad regional.

Representamos a la región norte por región 1 (r_1), el centro-norte por región 2 (r_2), el centro por región 3 (r_3) y la región sur por región 4 (r_4). En la Figura 3.2 se pueden visualizar las regiones de México.

Figura 3.2. *Regiones de México, de acuerdo a Banco de México (2011-2021)*



Fuente: Elaboración propia

Para la estimación a nivel regional, se toman a los entrevistados que radican en la misma región donde radicaba su PP. Lo que permite interpretar los resultados como indicadores de movilidad propios de las regiones.

Con los individuos que no migran se realiza una partición por regiones. Finalmente, cada región se analiza en el mismo sentido que se analizó a nivel nacional, es decir, bajo un modelo fijo y un modelo promedio, tanto para la estimación de EII como la AIR.

Las estimaciones se encuentran en la Tabla D.2 para la selección de gemelos sintéticos con el criterio A , y en la Tabla D.3 para la selección de gemelos sintéticos con el criterio B , ambas tablas se encuentran en el apéndice D. La discusión de los resultados se encuentra en el capítulo 5.

3.4. Interpretación: EII, AIR

Para simplificar la interpretación de EII, se toma la ecuación (3.1) sin controles para dos individuos distintos $\{i, j\}$,

$$\log(y_i^a) = \beta_0 + \log(y_i^p)\beta + v_i$$

$$\log(y_j^a) = \beta_0 + \log(y_j^p)\beta + v_j$$

Si se toma la diferencia de estas ecuaciones y posteriormente el promedio

$$\mathbb{E}(\log(y_i^a) - \log(y_j^a)) = \beta(\log(y_i^p) - \log(y_j^p)) \quad (3.4)$$

Se tiene que la diferencia porcentual entre los ingresos de dos individuos en la generación actual es una proporción β de la diferencia porcentual de sus PP en la generación previa, entonces, la movilidad relativa es menor si β es grande, y mayor si β es pequeña.

Ahora, de la ecuación (3.3) se toma el valor esperado del rango percentil de un individuo en la generación actual cuyo PP se encontraba en el rango percentil más alto de su distribución nacional

$$\mathbb{E}(R(y_i^a))_{100} = \mathbb{E}(R(y_i^a)|_{\rho=100}) = \alpha + 100\rho$$

De forma similar, se toma el valor esperado del rango percentil de un individuo en la generación actual cuyo PP se encontraba en el rango percentil más bajo de su distribución

nacional

$$\mathbb{E} \left(R(y_j^a) \right)_0 = \mathbb{E} \left(R(y_j^a) \Big|_{\rho=0} \right) = \alpha$$

Al tomar la diferencia, se tiene:

$$\mathbb{E} \left(R(y_i^a) \right)_{100} - \mathbb{E} \left(R(y_j^a) \right)_0 = 100\rho \quad (3.5)$$

Así, la movilidad relativa ρ , es una proporción de la diferencia entre el rango percentil esperado de un individuo cuyo PP se encontraba en la parte superior de la distribución nacional y el de un individuo cuyo PP se encontraba en la parte inferior de la distribución nacional.

Adicionalmente, se define la *Movilidad absoluta hacia adelante* (MAA), como el rango percentil esperado en la generación actual de individuos cuyos PP estaban en promedio debajo de la mediana de su distribución, lo cual es equivalente a (Chetty y col., 2014)

$$\text{MAA} = \alpha + 25\rho \quad (3.6)$$

De acuerdo a Delajara y Graña (2018), más que la elasticidad, el coeficiente de correlación es el objeto de estudio principal de la movilidad. En ese sentido, el coeficiente de correlación ρ ofrece una ventaja sobre β , pues la EII es una estimación de la correlación entre las variables analizadas.

3.5. Muestreo boopstrap

Para realizar inferencia se emplea un esquema bootstrap. Esto es necesario porque la estimación de la EII y la AIR implican primero estimar los ingresos de los individuos en ambas generaciones.

El primer paso para realizar la estimación es filtrar a los entrevistados de E17 con las características necesarias para encontrar a sus gemelos sintéticos y estimar los ingresos de su PP.¹³ Posteriormente se buscan a los gemelos sintéticos en las encuestas E16 y E18. Se toman las encuestas E94, E96, E98, y se filtran a los PPS que satisfagan las características

¹³ Las características se precisan en el capítulo 4.

reportadas por los entrevistados de E17.

Para los PPS se toma una muestra con reemplazo del tamaño total de la muestra y se calcula la primer etapa del método de MC2E2M. Luego, se toma una muestra con reemplazo de los gemelos sintéticos para cada entrevistado y se calcula el promedio de sus ingresos, este se asigna al entrevistado como ingreso estimado. De esta base de datos de padres sintéticos e ingresos estimados se toma una muestra con reemplazo del mismo tamaño que el número de entrevistados de la E17 y se obtiene una estimación de EII y AIR. Este procedimiento se realiza 100 veces, en donde cada resultado esta determinado por los muestreos realizados.

El método de muestreo con reemplazo, conocido como bootstrap no paramétrico, permite construir una distribución empírica del estimador de interés y así realizar inferencia estadística, en este caso se utiliza para la estimación de los errores estándar de los indicadores EII y AIR, para posteriormente realizar la prueba t para dichos parámetros.

Para calcular el error estándar de los estimadores, se toma $\hat{\theta}_1, \dots, \hat{\theta}_B$, donde $\hat{\theta} \in \{\hat{\beta}, \hat{\rho}, \hat{\alpha}\}$. El error estándar del estimador es la raíz cuadrada de $s_{\hat{\theta}, B}^2$:

$$s_{\hat{\theta}, B}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \bar{\hat{\theta}})^2, \text{ donde } \bar{\hat{\theta}} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b$$

Todos los errores estándar reportados se calcularon con esta metodología, como sugiere la literatura (Báez, 2021). Debido a los requerimientos de cómputo se tomó $B = 100$. Algunas pruebas se ejecutaron con $B = 500$ y $B = 1000$ pero los resultados no tenían cambios significativos.

En este capítulo se dan los detalles sobre los datos utilizados.

Los datos para el análisis de movilidad intergeneracional se obtuvieron de las encuestas ESRU - EMOVI - 2017 (E17) y de las encuestas ENIGH 1994 (E94), 1996 (E96), 1998 (E98), 2016 (E16) y 2018 (E18). La información de la encuesta E17 se utiliza en dos momentos, en un primer momento junto con las encuestas E94, E96 y E98, para obtener la estimación de los ingresos del proveedor principal (PP). En un segundo momento para cruzar información con las encuestas E16 y E18 para la estimación de los ingresos actuales.

4.1. Encuesta ESRU - EMOVI

La información para construir los ingresos del y^p y los ingresos actuales y^a , se obtienen de la encuesta ESRU - EMOVI (encuesta de la fundación Espinosa Rugarcía de Movilidad Social en México), en particular se selecciona el levantamiento 2017. Esta encuesta tiene como finalidad principal medir la movilidad social intergeneracional. Cuenta con 17,665 entrevistas y es representativa de hombres y mujeres entre 25 y 64 años a nivel nacional (CEEY, 2018).

Con la finalidad de recabar información sobre los PP, la encuesta contiene secciones donde se pregunta sobre el contexto económico del entrevistado a los 14 años de edad. Mientras que otras secciones se refieren al contexto económico actual del entrevistado.

Para realizar la estimación de los ingresos actuales y pasados, es necesario seleccionar a los individuos que proporcionen tanto información personal como la de su PP, por tal motivo,

las características que debe reportar el entrevistado son las siguientes:

- Vivir en un hogar donde se comparte el gasto de alimentos
- Edad entre 25 y 50 años (para incrementar el tamaño de muestra)
- Contar con trabajo
- No ser estudiante
- Nivel de educación máxima
- Ocupación
- Vivir en un hogar donde sola una persona aporta los ingresos del hogar
- Intervalo de ingreso del hogar
- Los antecedentes de educación, ocupación, sexo, lugar de residencia y edad de su proveedor principal (padre o madre) a los 14 de años de edad

La última característica será utilizada para construir el ingreso del PP, mientras que las demás características servirán para construir el ingreso actual puntual, debido a que el registro de ingreso del entrevistado se encuentra a nivel de intervalo.

Los entrevistados con estas características son 1503. Como se comentó anteriormente, el ingreso de los hogares a los que pertenece el individuo en 2017 se reporta por intervalos de ingreso, es decir, el entrevistado indicó cuantas personas aportaban ingresos al hogar y en qué intervalo estaba la suma de los ingresos. Los intervalos de ingreso tienen la definición mostrada en la Tabla 4.1.

Tabla 4.1. *Intervalos de ingreso*

	Intervalo
Menor a un salario mínimo	(0, \$2399)
Igual a un salario mínimo	\$2400
Más de un salario mínimo hasta 2	(\$2401, \$4800]
Más de 2 salarios mínimos hasta 3	(\$4801, \$7200]
Más de 3 salarios mínimos hasta 5	(\$7201, \$12000]
Más de 5 salarios mínimos hasta 10	(\$12001, \$24000]
Más de 10 salarios mínimos	(\$24001, o más)

Fuente: Elaboración propia

Para aplicar el método de MC2E2M es necesario estimar el ingreso para cada entrevistado i . La estrategia que se sigue es la siguiente:

1. Se toma el sexo, la edad, la ocupación, la educación, y la región de residencia del entrevistado
2. Se buscan individuos en las encuestas E16 o E18 (que llamaremos *gemelos sintéticos*), que compartan dichas características con un cierto grado de flexibilidad (criterio A o criterio B). Estas encuestas son las más cercanas a E17, tomar ambas aumenta la posibilidad de encontrar gemelos sintéticos de una mayor cantidad de entrevistados
3. Cada individuo i tendrá un conjunto de gemelos sintéticos asignado $Y_{it'}^a$, con $t' \in \{2016, 2018\}$ dependiendo en donde se ubiquen sus gemelos sintéticos.

Lo anterior permite tener distintas posibilidades de asignación puntual de ingresos para cada entrevistado, en particular, se asigno el valor promedio de los ingresos de los gemelos sintéticos de cada entrevistado. La descripción sobre la elección de esta metodología se encuentra en la sección 4.1.1.

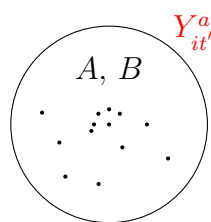
Para determinar los gemelos sintéticos usados para estimar los ingresos de los individuos en la E17, se siguen dos criterios:

- **Criterio A.** Se buscan *gemelos sintéticos* que tengan dos años a lo mas de diferencia con el entrevistado, y que la ocupación coincida de lo particular a lo general,¹⁴ las demás características seleccionadas se buscan se satisfagan de forma exacta.
- **Criterio B.** A diferencia del criterio A , la edad se busca de forma exacta, además se restringe que los gemelos sintéticos tengan en la E16 y E18 ingresos dentro del intervalo reportado en E17.

El criterio A permite que el entrevistado tenga mayor cantidad de gemelos sintéticos, mientras que el criterio B asigna menos gemelos sintéticos pero las especificaciones son más precisas. La Figura 4.1 muestra un esquema de la construcción del conjunto de gemelos sintéticos. La cantidad de gemelos sintéticos asignada van desde un gemelo hasta individuos con 176 gemelos sintéticos.

¹⁴ Los códigos de ocupación tienen 4 dígitos, un dígito para cada subclasificación, una coincidencia exacta implica la coincidencia de los 4 dígitos, la coincidencia más general implica la coincidencia del primer dígito.

Figura 4.1. Para cada individuo se construyen gemelos sintéticos que tengan las características del criterio A o el criterio B en la base del año t' .



Fuente: Elaboración propia

Con el criterio A (criterio B) se lograron construir los ingresos de 1486 (1422) entrevistados para el análisis de movilidad nacional. Para el análisis regional, la muestra es de 1365 (1305). La Tabla 4.2 muestra la composición de las bases de datos. Con el criterio A (criterio B) a nivel regional, se cuentan con: 232 (225) entrevistados que residen en la región 1, 374 (357) en la región 2, 500 (479) en la región 3 y 259 (245) en la región 4, que comparten residencia con su PP.

Tabla 4.2. Base de datos construida con los criterios A y B

	Criterio A	Criterio B
Hombres	885	848
Mujeres	601	574
Promedio edad entrevistado	36.5	36.4
PP Hombres	1,219	1163
PP Mujeres	267	259
PP edad mínima	16	16
PP edad máxima	79	79
PP edad promedio en 2017	64	64

Fuente: Elaboración propia

La edad mínima y máxima del PP se refiere al momento que el entrevistado tenía 14 años. La edad promedio de los entrevistados nos permite ubicar la encuesta principal para la construcción del ingreso del PP en 1995. Como no existe levantamiento de ENIGH en 1995, se tomaron E94, E96 y E98, con los PPS en el rango de edad entre 25 y 60 años.

Para ambos criterios de selección de gemelos sintéticos, se tienen entrevistados seleccionados de todas las edades elegidas, en todos los estados del país, y en todos los intervalos de ingreso propuestos en E17,. Los valores específicos de esta información se encuentran en el apéndice A.

Se observa, con ambos criterios de construcción, que la muestra es pequeña,¹⁵ aún considerando a hombres y mujeres, es por ello que en las estrategias de estimación se controla por sexo de los individuos.

4.1.1. Selección de variables para la estimación de ingresos

Dado que en la encuesta E17 solo se conoce el intervalo de ingreso del hogar de los individuos, se propone una metodología para seleccionar las características que permitan determinar quiénes son los potenciales gemelos sintéticos en la E16 y E18, que sea consistente con las características de los individuos en la E17.

La idea es seleccionar un conjunto de características que nos permita predecir en qué intervalo de ingreso se encuentra cada individuo, de acuerdo con su autoreporte en la E17, con la mayor precisión posible. Para esto se emplean diversas técnicas de aprendizaje automático (AA), como máquinas de soporte vectorial, vecinos más cercanos, regresión logística, redes neuronales, entre otras.

Después de estimar con distintas especificaciones de los modelos, con distinta cantidad de variables independientes, y seleccionando variables bajo distintos criterios. Se obtuvo que utilizando las variables independientes: región, sexo, edad, educación y ocupación, en promedio la predicción del intervalo de ingreso era correcta en un 33 % de las veces.¹⁶

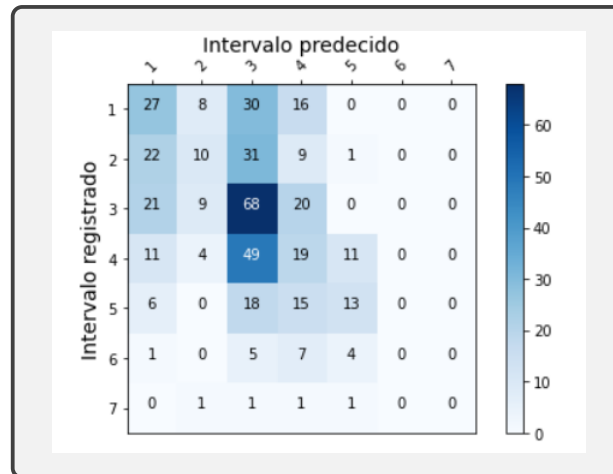
La Figura 4.2 muestra la precisión de la clasificación con las variables mencionadas, para una especificación (variar los parámetros del algoritmo produce distintas aproximaciones) de las máquinas de soporte vectorial. Para ejemplificar el cálculo de la precisión del algoritmo obsérvese la fila 3 de la Figura 4.2. Se tomaron 118 individuos cuyo intervalo de ingreso reportado fue el tercero, mientras que el algoritmo predijo de forma correcta el intervalo de ingreso de 68 individuos, con lo que se obtiene una precisión del 57.6 %.¹⁷

¹⁵ Esto es una de las motivaciones para hacer uso de muestreo con reemplazo en la estimación de los errores usando bootstrap.

¹⁶ Con una muestra independiente de entrevistados se observaron las características mencionadas y en el 33 % de las veces, el algoritmo clasificó de acuerdo al intervalo reportado por el entrevistado.

¹⁷ Los individuos que se utilizaron para encontrar la precisión son diferentes a los utilizados para ajustar los parámetros del algoritmo.

Figura 4.2. Precisión de las máquinas de soporte vectorial.



Fuente: Elaboración propia

Se considera que estos resultados son consecuencia de la heterogeneidad de los individuos, pues dentro de la muestra, hay individuos con las mismas características que reportan diferente intervalo de ingreso. Por ejemplo, existen individuos que comparten la característica de vehículos en propiedad, pero tienen diferente registro de intervalo de ingreso, y lo mismo ocurre con otras variables (por ejemplo: educación).

La idea de la solución propuesta es tener pocas variables de selección,¹⁸ y agrupar a individuos (de otras encuestas) que tengan estas características (*gemelos sintéticos*) pero con un ingreso puntual asignado. Además la realización de muestreo con reemplazo permite mayor variación en las posibles asignaciones de ingreso.¹⁹

La motivación en la elección de los criterios surge de los resultados de los algoritmos de AA, pues como se puede observar en la Figura 4.2, tener las características mencionadas, puede subestimar o sobrestimar el intervalo de ingreso. De acuerdo a la flexibilidad de cada criterio, se tienen que: el criterio *A* tiene una mayor varianza en la asignación de ingreso, mientras el criterio *B* reduce la varianza en la asignación. La implementación y resultado de los algoritmos se puede encontrar en el apéndice en línea.²⁰

¹⁸ El tener bastantes variables no aumenta en gran medida la precisión (aumentaba hasta un 40 %) de la clasificación y reduce el espacio de búsqueda.

¹⁹ Después de realizar un muestreo con reemplazo de los gemelos sintéticos, la asignación puntual de ingreso es el promedio de los gemelos sintéticos encontrados, como se indica en el capítulo 3.

²⁰ La dirección es la siguiente: <https://rafneta.github.io/TesinaCIDEMERMM/index.html>

4.2. Encuestas ENIGH

La Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) tiene como objetivo proporcionar un panorama estadístico del comportamiento de los ingresos y gastos de los hogares en cuanto a su monto, procedencia y distribución; adicionalmente, ofrece información sobre las características ocupacionales y sociodemográficas de los integrantes del hogar, así como las características de la infraestructura de la vivienda y el equipamiento del hogar (INEGI-ENIGH, 1994, 1996, 1998, 2016, 2018).

Como se indicó en la sección anterior, se seleccionaron los levantamientos E94, E96 y E98, que servirán para reducir la DA comentada en el capítulo 3. En estas encuestas se seleccionaron individuos que reporten las siguientes características

- Edad entre 25 y 60 años
- Sexo
- Lugar de residencia
- Ocupación
- Ingresos
- Nivel de educación
- Hijos en el rango de edad específico

con la finalidad de empatarlas con las reportadas por los entrevistados seleccionados de la E17. Para tomar en cuenta que la edad promedio de los entrevistados en E17 era de 14 años en 1995, se seleccionaron aquellos individuos con hijos menores que 13, 16 y 18 años para las encuestas E94, E96 y E98 respectivamente.

Se codificaron a las ocupaciones en 19 categorías (O_1 a O_{19}), la educación en cinco categorías (E_1 a E_5) y la región de residencia en las cuatro categorías descritas anteriormente (r_1 a r_4). El apéndice C presenta el detalle de esta agrupación. Se descartaron las fuentes de ingreso que no son frecuentes, como aguinaldos, primas, bonos, etc.

La Tabla 4.3 presenta un resumen de los ingresos y la proporción de hombres y mujeres que son los PP en las muestras de E94, E96 y E98. Los ingresos se deflactaron con el Índice Nacional de Precios al Consumidor de Banco de México a enero de 2017.

Tabla 4.3. Ingresos del hogar

Encuesta	<i>N</i>	Hombres	Mujeres	Promedio ingreso
E94	4,694	4,362	332	1,588.144 [7.370]
E96	5,980	5,510	470	2,078.353 [7.639]
E98	4,729	4,284	445	3,309.484 [8.105]

Nota: Desviación estándar entre corchetes.

Fuente: Elaboración propia

La distribución de los ingresos, el promedio de ingresos por edad, tanto nacional como por región se muestran en el apéndice B. En estas muestras se puede apreciar una variación de los ingresos con la edad. Por esta razón, en las estrategias de estimación se controla por un polinomio de segundo grado dependiente de la edad. Además, existe una clara variación de los ingresos por región, es por ello que un análisis regional es necesario.

Para la estimación de los modelos AIR, es necesario recuperar la distribución nacional de ingresos. Dicha distribución se pueden observar en la Figura B.18 del apéndice B.

Para las encuestas E16 y E18, no se recuperan a todos los individuos, salvo aquellos que son gemelos sintéticos de los encuestados en E17. Sin embargo, al estimar el AIR, es necesario la ubicación del percentil nacional de los ingresos, por tal motivo es necesario elaborar las densidades de ingresos de las encuestas completas, nacionales y a nivel región.

Tomando en cuenta que los gemelos sintéticos se encuentran en E16 y E18, también se construye la densidad de la unión de los individuos en ambas muestras.²¹ Para el cálculo de los ingresos, se descartaron aquellos que no son frecuente, como aguinaldos, primas, bonos, etc. Los ingresos se deflactaron con el Índice Nacional de Precios al Consumidor de Banco de México a enero de 2017. La distribución de ingresos se pueden observar en la Figura B.18 del apéndice B.

²¹ Las densidades estimadas usando cada muestra por separado no difieren en gran medida. Entonces tomando la densidad de la unión de ambas muestras ofrecen una buena aproximación de lo que sucedía en 2017.

Resultados y discusión

La presentación de resultados versa sobre la interpretación de las estimaciones de los parámetros de los modelos (ME_{np}) , (MR_{np}) , y la versión regional de estos, pues si bien los modelos fijos son la base para el cálculo, resultan más propensos a la DA.

5.1. Elección de estimación

Con el fin que la interpretación de resultados esté menos influenciada por la DA, el análisis resultados se realizará sobre los modelos promedio, tanto a nivel nacional como regional. Cabe recordar que en el modelo promedio, la estimación de ingreso asignada al PP es el promedio de las estimaciones obtenidas con E94, E96 y E98 (modelos fijos). Los resultados de las estimaciones con los modelos fijos se pueden encontrar en las Tablas D.2, D.3 y D.1, del apéndice D.

Debido a que los ingresos actuales se tuvieron que estimar, los resultados de movilidad difieren substancialmente dependiendo si la estimación se realizó con el criterio *A* o el criterio *B*. Como se definió en la sección 4.1, la principal característica de selección de gemelos sintéticos con el criterio *A* es que el rango de ingreso es libre. En cambio, con el criterio *B* se restringe a que el gemelo sintético tenga ingreso en el mismo intervalo reportado por el entrevistado de E17. La Tabla 5.1 muestra los resultado nacionales para ambos criterios.

Con ambos criterios se obtienen valores de estimación por debajo de los valores reportados en Báez (2021) y Campos-Vázquez y col. (2020). En Báez (2021), se realiza la estimación aplicando el método de MC2E2M, donde se obtiene una estimación de la EII de 0.79 y una AIR de 0.315, como base actual se utiliza ESRU - EMOVI - 2011 y como bases de pseudo PP, ENUE (1987 a 1991). Si bien las muestras empleadas en cada estudio

Tabla 5.1. *Estimación de movilidad nacional*

	Criterio A	Criterio B
$\hat{\beta}$ (EII)	0.247*** (0.038)	0.359*** (0.064)
$\hat{\rho}$ (AIR)	0.226*** (0.029)	0.262*** (0.034)
$\hat{\alpha}$	45.363*** (1.598)	28.937*** (1.873)

Nota: *** $p < 0.01$.

Fuente: *Elaboración propia*

son distintas y hay una diferencia en promedio de 6 años en las generaciones analizadas, la diferencia es que en este trabajo se estimaron los ingresos actuales, además que se tiene una muestra actual (E17) con relativamente pocas observaciones.

En Campos-Vázquez y col. (2020), se utiliza un modelo diferente al utilizado en este trabajo, donde se obtiene una EII en el rango de $[0.451, 0.5]$ (dependiendo de la especificación), y una AIR de 0.5. A diferencia del presente trabajo, Campos-Vázquez y col. (2020) emplean censos de población (1990, 1995, 2000, 2010, y 2015) con observaciones reportadas que rondan los dos millones y medio de observaciones. A pesar de las diferencias, se encuentra similitud en los años de las muestras utilizadas con las de este trabajo, además que bajo ciertas especificaciones que reportan estos autores, las estimaciones obtenidas de la EII son cercanas a las estimaciones reportadas en la Tabla 5.1.

Respecto a la movilidad regional, la Tabla 5.2 muestra los resultados obtenidos de las estimaciones. En Báez (2021), los valores estimados de la EII para las regiones son: norte $\hat{\beta} = 0.371$, $\hat{\rho} = 0.130$; centro-norte $\hat{\beta} = 0.377$, $\hat{\rho} = 0.163$, centro $\hat{\beta} = 0.627$, $\hat{\rho} = 0.308$, sur $\hat{\beta} = 0.974$, $\hat{\rho} = 0.403$. Mientras que en Campos-Vázquez y col. (2020), solo se reporta la EII de las regiones, cuyos valores son: norte $\hat{\beta} = 0.434$, centro-norte $\hat{\beta} = 0.496$, centro $\hat{\beta} = 0.515$, sur $\hat{\beta} = 0.534$.

Tabla 5.2. Estimación regional de movilidad

<i>Región de México</i>				
<i>Panel A: Criterio A</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\rho}$	0.254** (0.113)	0.295*** (0.076)	0.216*** (0.068)	0.343*** (0.080)
$\hat{\rho}$	0.216*** (0.073)	0.271*** (0.057)	0.208*** (0.048)	0.328*** (0.064)
$\hat{\alpha}$	50.513*** (4.642)	42.815*** (3.193)	43.817*** (2.835)	38.431*** (2.968)
<i>Panel B: Criterio B</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\beta}$	0.325** (0.157)	0.385*** (0.128)	0.310*** (0.114)	0.236 (0.154)
$\hat{\rho}$	0.198* (0.113)	0.274*** (0.080)	0.255*** (0.067)	0.164** (0.076)
$\hat{\alpha}$	37.525*** (6.708)	27.247*** (4.391)	30.472*** (3.942)	23.971*** (3.244)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Fuente: Elaboración propia

5.2. Movilidad nacional

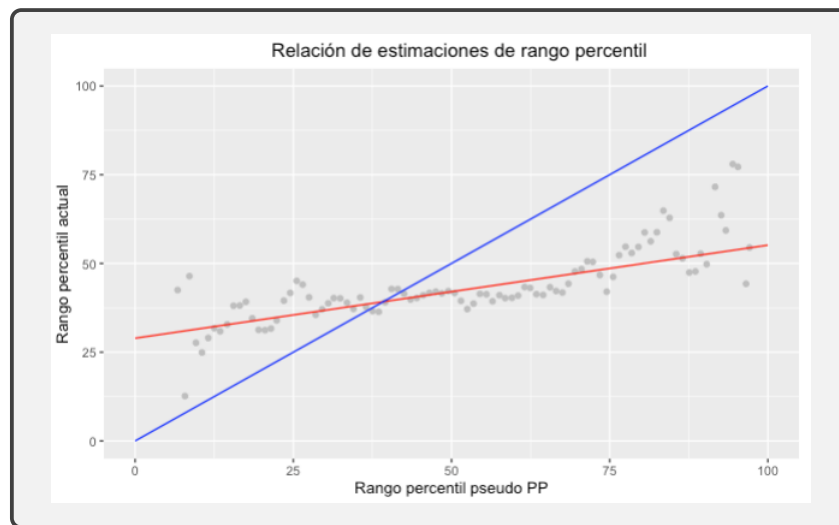
Una EEI estimada de 0.359 como la presentada en la Tabla 5.1, implica que aproximadamente el 36 % de las diferencias relativas porcentuales es transmitida entre las generaciones, esto habla de existencia de movilidad alta. Sin embargo, una interpretación más completa de los patrones de movilidad se auxilia de los resultados del modelo AIR.

La estimación de la AIR para la muestra nacional es $\hat{\rho} = 0.262$ y $\hat{\alpha} = 28.937$. Esto implica que la diferencia promedio del rango percentil de un individuo que proviene de un PP que se encontraba en la parte superior de la distribución, respecto a un individuo que proviene de un PP que se encontraba en la parte inferior de la distribución es pequeña, entonces existe movilidad.

Para entender las implicaciones de los resultados de la estimación del modelo AIR, la Figura 5.1 representa la relación del rango percentil de la generación actual con la del correspondiente PPS. Los puntos grises de la Figura 5.1 se obtienen al dividir ambos rangos

percentiles en cien intervalos de igual longitud y promediando los valores de los ingresos que caen dentro del intervalo correspondiente (Chetty y col., 2014). La línea roja muestra el ajuste de un modelo lineal. Se espera que, en promedio, la relación entre rangos percentiles sea lineal. Para el caso de los datos analizados, se observa una alta varianza en los rangos más altos y más bajos de los PPS, lo cual es capturado en las estimaciones de movilidad.

Figura 5.1. *Relación entre el rango percentil promedio actual estimado y PP.*

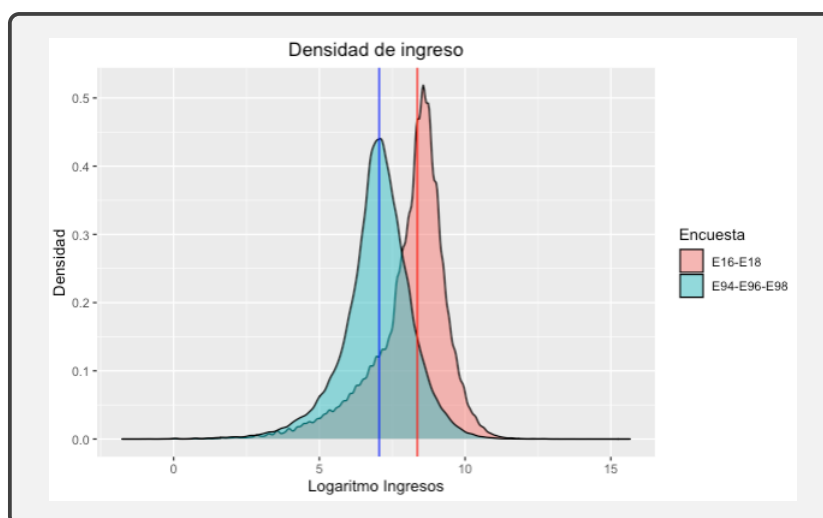


Fuente: Elaboración propia

En la Figura 5.1 la línea azul es la recta identidad y representa una situación en la que el rango de ingresos actual y de origen es el mismo. El cruce entre la línea roja y la azul (en este caso, 38.4 % del rango percentil de los PPS) indica que los individuos con un PPS con un rango por debajo de dicho punto, hoy se encuentran en una situación comparativamente mejor, con una situación en un rango percentil más alto que el de su correspondiente PPS. Lo contrario ocurre con los puntos a la derecha del cruce. Así, en promedio los individuos están relativamente más cerca y esto se registra en los parámetros como movilidad.

Para verificar esta última observación, en la Figura 5.2 se muestran las distribuciones de los ingresos nacionales de E94-E96-E98 y E16-E18.

Figura 5.2. *Densidad de ingreso nacional*



Fuente: *Elaboración propia*

Las líneas verticales muestran la ubicación de la mediana nacional en las respectivas muestras. La cola izquierda de la distribución de ingresos E16-E18 es ancha, es decir, existe más población con niveles bajos de ingreso. Esto no sucede en la distribución de ingresos de E94-E96-E98. Este comportamiento es capturado en los resultados obtenidos.

Por otro lado, los resultados anteriores también nos permiten decir algo sobre la movilidad absoluta. La *MAA* es de 34.821%, este es el rango percentil esperado en la generación actual de individuos cuyos PPS estaban, en promedio, debajo de la mediana de su distribución de ingresos.

Finalmente, otra forma de observar los patrones de movilidad consiste en calcular la probabilidad de que los ingresos actuales de un individuo estén en el quintil q_i , $i \in \{1, \dots, 5\}$, dado que los ingresos de su PPS se encontraban en el quintil Q_j , $j \in \{1, \dots, 5\}$. La Tabla 5.3 muestra estos cálculos.

Si calculamos la probabilidad acumulada de que los ingresos actuales estén en los quintiles q_2 y q_3 , dado que los de PPS estaban en Q_1 , se tiene que $P(q = q_2, q = q_3 | Q = Q_1) > P(q = q_1 | Q = Q_1)$. Es por eso que en la Figura 5.1 el ajuste de la AIR está por arriba de la identidad en la región que corresponde a Q_1 . En esta misma figura se observa que el promedio de los ingresos de individuos actuales que

Tabla 5.3. Probabilidad (%) de cambiar de quintil

Quintil PPS	Quintil actual				
	q_1	q_2	q_3	q_4	q_5
Q_1	32.6	28.2	21.6	13.3	4.3
Q_2	31.6	23.9	23.0	14.4	7.2
Q_3	25.5	23.7	24.1	18.6	8.2
Q_4	21.0	21.2	25.0	21.9	10.9
Q_5	14.0	17.2	18.4	20.8	29.6

Fuente: Elaboración propia

proviene de PPS en el Q_1 , no supera la mitad del q_3 .²² Es decir, la situación de mejores ingresos promedio, de individuos actuales que provienen de PPS ubicados en Q_1 , está acotada.

Bajo un ejercicio similar al anterior, se puede concluir sobre el comportamiento del ajuste del modelo AIR en la región que corresponda a los Q_j restantes.

²² Puntos en gris en la región indicada.

5.3. Movilidad regional

La representatividad de la E17 permite realizar los análisis anteriores para cuatro grandes regiones de México. Los resultados de estimar los parámetros de los modelos de EII y AIR a nivel regional se presentan en la Tabla 5.4. La región más móvil es el sur, seguida del norte, en tercer lugar el centro y por último el centro-norte.

Los valores de $\hat{\beta}$ indican que hay movilidad en todas las regiones, pero la estimación en la región sur no es precisa. Al igual que el caso nacional, los resultados de la estimación del modelo AIR nos permiten hacer una exploración más detallada de los patrones de movilidad.

Tabla 5.4. *Estimación regional de movilidad*

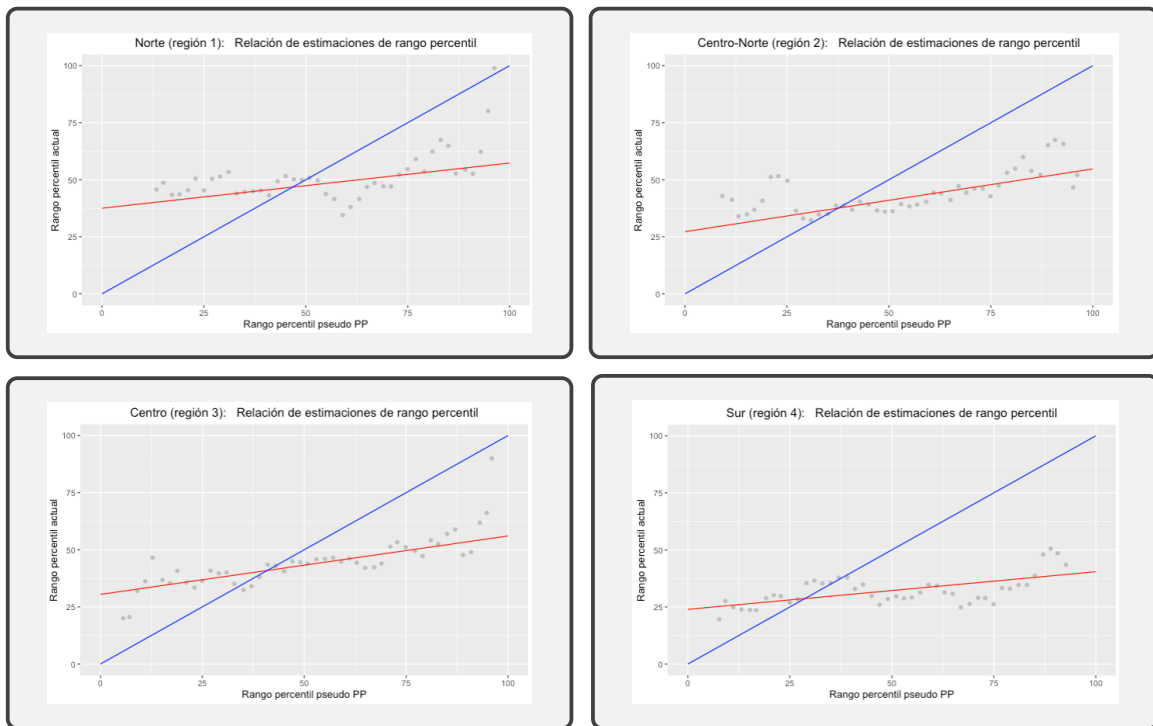
	<i>Región de México</i>			
	norte	centro-norte	centro	sur
Movilidad relativa				
$\hat{\beta}$	0.325**	0.385***	0.310***	0.236
$\hat{\rho}$	0.198*	0.274***	0.255***	0.164**
$\hat{\alpha}$	37.525***	27.247***	30.472***	23.971***
Movilidad absoluta				
<i>MAA</i>	42.475	34.097	36.847	28.071

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Fuente: *Elaboración propia*

La construcción de la Figura 5.3, al igual que el caso nacional, se realizó siguiendo a Chetty y col. (2014). La recta en rojo es el modelo estimado de AIR para la región correspondiente, y la recta en azul es la recta identidad.

Figura 5.3. Relación entre de rango percentil promedio actual y PPS.



Fuente: Elaboración propia

Al igual que el caso nacional, la movilidad capturada por la AIR se debe a que los individuos que provienen de PPS ubicados en percentiles altos, en promedio estarán ocupando posiciones menos favorables en la distribución actual. Mientras individuos actuales que provienen de PPS ubicados en percentiles bajos, en promedio están un poco mejor. Lo que tiene como consecuencia que la distancia promedio entre su ubicación actual $\hat{\rho}$, sea pequeña.

El fenómeno es parecido en todas las regiones, pero en el norte y el sur existen comportamientos peculiares, que se describirán a continuación.

En la región norte, todos aquellos individuos que provienen de PPS que estaban por debajo del 46.789% (intersección de la recta roja y azul en la región correspondiente de la Figura 5.3), en su distribución de ingreso nacional, ahora están en una situación *mejor* (la recta en rojo esta por arriba de la recta azul, hasta antes del punto de intersección).

Pero todos aquellos individuos que provienen de PPS que estaban ubicados por arriba del 46.789% en su distribución nacional, ahora se encuentran, en promedio, en una situación menos favorable. En la región centro-norte, centro y sur, el comportamiento es similar, pero

los puntos de referencia son: 32.530 %, 40.902 % y 28.637 %, respectivamente.

Finalmente, si se calcula la proporción de ingresos estimados de individuos actuales que están dentro de un quintil de ingreso q_i , $i \in \{1, \dots, 5\}$, respecto del quintil en el cual se encontraba su PPS Q_j , $j \in \{1, \dots, 5\}$, se obtienen los resultados reportados en las Tablas 5.5, 5.6, 5.7 y 5.8.

Tabla 5.5. *Región norte: Probabilidad (%) de cambiar de quintil*

Quintil PPS	Quintil actual				
	q_1	q_2	q_3	q_4	q_5
Q_1	24.0	12.6	29.9	33.5	0
Q_2	19.8	15.9	33.5	18.0	12.8
Q_3	18.9	17.6	26.1	25.4	12.0
Q_4	19.1	17.8	26.0	23.4	13.7
Q_5	13.0	12.0	16.4	23.8	34.7

Fuente: Elaboración propia

Tabla 5.6. *Región centro-norte: Probabilidad (%) de cambiar de quintil*

Quintil PPS	Quintil actual				
	q_1	q_2	q_3	q_4	q_5
Q_1	22.7	43.2	17.2	4.4	12.6
Q_2	35.7	21.6	20.8	13.8	8.1
Q_3	28.0	26.6	23.8	13.1	8.5
Q_4	22.8	21.5	21.3	21.2	13.2
Q_5	9.6	15.3	20.8	28.8	25.6

Fuente: Elaboración propia

Tabla 5.7. *Región centro: Probabilidad (%) de cambiar de quintil*

Quintil PPS	Quintil actual				
	q_1	q_2	q_3	q_4	q_5
Q_1	26.8	23.6	32.0	15.6	2.0
Q_2	28.0	26.5	25.7	15.2	4.5
Q_3	20.5	21.4	26.7	21.1	10.2
Q_4	17.8	20.0	28.7	23.7	9.7
Q_5	11.5	14.6	28.9	18.8	26.1

Fuente: Elaboración propia

Estos resultados, al igual que el caso nacional, son una forma de cuantificar por quintiles los resultados del modelo AIR. En general estos resultados se deben de interpretar con cautela, debido a que la muestra con la que se obtuvieron es pequeña, al final de este capítulo se comentará más al respecto.

Tabla 5.8. Región sur: Probabilidad (%) de cambiar de quintil

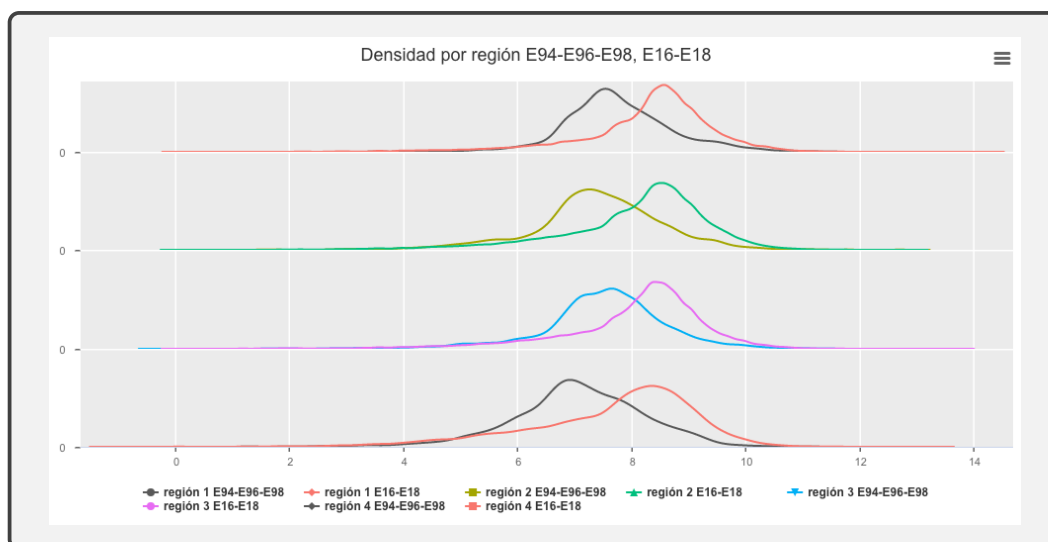
Quintil PPS	Quintil actual				
	q ₁	q ₂	q ₃	q ₄	q ₅
Q1	52.3	23.7	16.4	7.5	0.2
Q2	32.5	33.6	20.5	10.8	2.5
Q3	42.2	28.8	13.5	11.9	3.6
Q4	36.0	30.6	22.0	8.8	2.7
Q5	24.6	32.6	9.6	10.8	22.5

Fuente: Elaboración propia

Lo que sí se puede precisar es el comportamiento cualitativo de las regiones. La región sur, tiene una mayor persistencia de los individuos que provienen de PPS en el primer quintil, y una alta probabilidad que individuos provenientes de PPS de quintiles superiores estén en una peor situación. Esto va cambiando conforme se recorre el país hacia la región norte, donde existe la mayor probabilidad de permanecer en el quintil más alto.

Esta tendencia de movilidad descendente puede ser explicada porque en las distribuciones regionales de E16-E18 existe una cola ancha izquierda, en comparación con las distribuciones regionales de E94-E96-E98. En la Figura 5.4, se muestran la densidades del logaritmo del ingreso por región.

Figura 5.4. Densidad del logaritmo del ingreso de la población por región, para E94-E96-E98 y E16-E18.



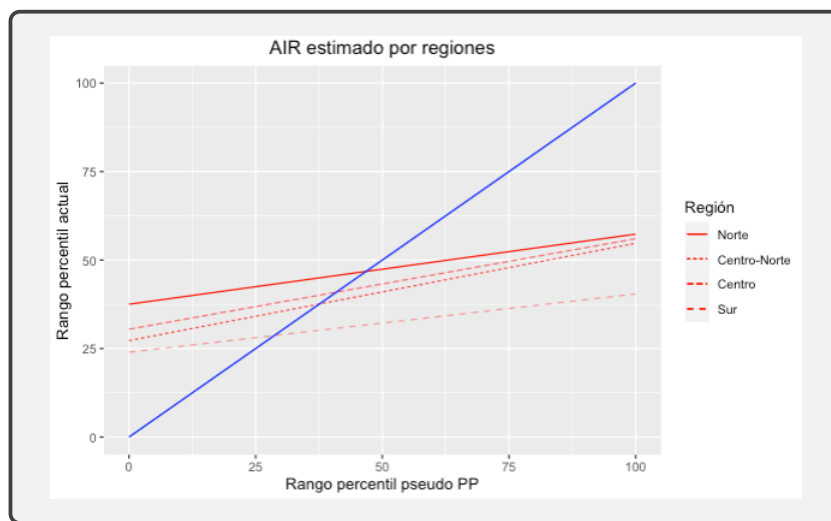
Fuente: Elaboración propia

En particular existe mayor masa en los quintiles inferiores en la región sur y en la región centro-norte. En ese sentido, las matrices de probabilidad reflejan ese comportamiento. Y

precisamente, esto tiene sentido con la movilidad absoluta ascendente, donde la región sur y la centro-norte son el último y penúltimo lugar regional respectivamente, como se observa en la Tabla 5.4.

Con esta información se concluye que la región sur es la que experimenta mayor movilidad hacia peores condiciones, mientras que la región norte es la menos perjudicada. La comparación se hace más evidente si se observa la Figura 5.5.

Figura 5.5. Modelo AIR estimado por regiones.



Fuente: Elaboración propia

Si se comparan estas dos regiones, se observa en la Figura 5.3, que en la región norte no se tienen registros de individuos cuyo origen sea un PPS en el primer decil, y los individuos que provienen de PPS en los percentiles más altos tienden a conservar su posición. Mientras que el sur, en promedio, hay una menor proporción de individuos actuales que provienen de PPS del decil más alto.

En los estudios de movilidad realizados para México, se presenta esta misma situación, en el sentido que el norte tiene una mejor situación de individuos actuales y el sur una peor situación (Báez, 2021; Campos-Vázquez y col., 2020; Delajara & Graña, 2018; Monroy-Gómez-Franco & Corak, 2020). En los resultados obtenidos se tiene que el sur presenta una mayor movilidad que el norte, pero esta movilidad no es hacia mejores condiciones.

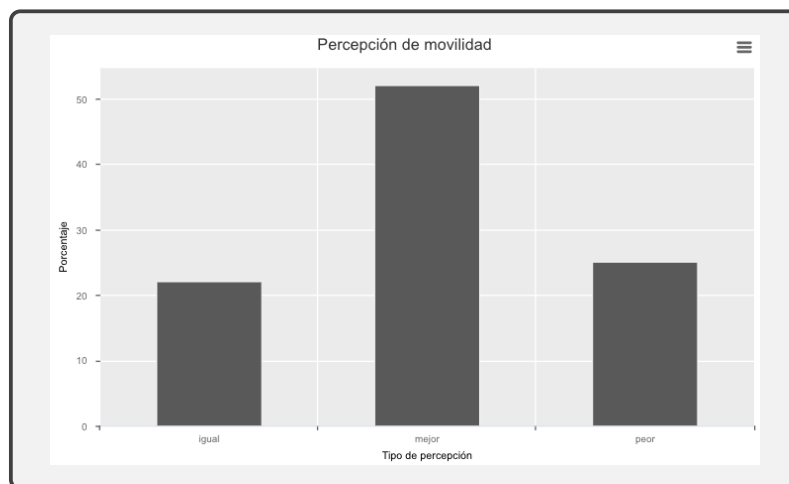
La región sur y la región norte, presentan la peculiaridad de que han sido las regiones con menores ingresos y con mayores ingresos, respectivamente. Esto se puede observar en la posición relativa de sus densidades de ingreso, como lo muestra la Figura 5.5. Incluso las Figuras B.12, B.13, y B.14, muestran que este comportamiento, en promedio, se mantiene por edad de los individuos. Adicionalmente la Figura E.3 muestra los niveles de ingreso promedio estimados para estas regiones, donde es clara la diferencia de nivel.

5.4. Movilidad percibida

Una forma de validar los resultados obtenidos es comparándolos con lo que los encuestados en la E17 reportan como su percepción de movilidad.

En la E17 se les pregunta a los encuestados: “Comparando su hogar con todos los hogares de México en este momento, ¿dónde pondría usted su hogar? (1. Más pobre, 10. Más rico). Y pensando en el hogar donde Ud. vivía a los 14 años, comparando ese hogar con todos los hogares de México de ese tiempo, ¿dónde pondría usted su hogar de ese entonces? (1. Más pobre, 10. Más rico)”. La Figura 5.6 muestra los resultados de dichas preguntas para todos los entrevistados, donde se han registrado las respuestas como, igual, mejor o peor, dependiendo si el número de respuesta de la situación actual respecto a la situación de origen, es igual, mayor, o menor, respectivamente.²³

Figura 5.6. *Percepción de movilidad de E17.*



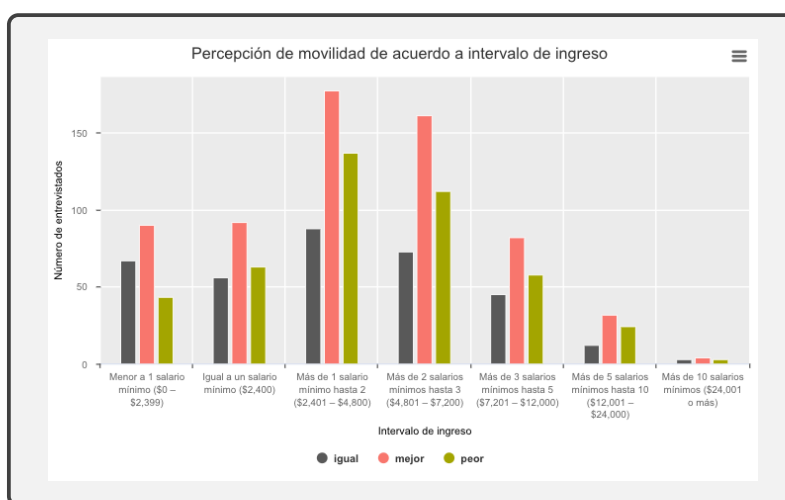
Fuente: Elaboración propia

²³ Los resultados son cualitativamente similares si se emplea solo la submuestra usada en las estimaciones de este trabajo.

Si bien es difícil que un entrevistado tenga una percepción nacional de su situación económica, esta estadística nos permite observar que la mayoría de los individuos en la E17, y en la muestra seleccionada, percibe una movilidad ascendente, en menor proporción movilidad descendente, y por ultimo una menor proporción no percibe movilidad.

Por otro lado, la mayor proporción de entrevistados que se encuentra en los quintiles inferiores, perciben experimentar una movilidad ascendente. Esto coincide con los resultados del modelo AIR. La Figura 5.7 muestra la movilidad percibida de acuerdo al grupo de ingreso reportado en la E17.

Figura 5.7. Percepción de movilidad muestra de E17 de acuerdo a intervalos de ingreso.

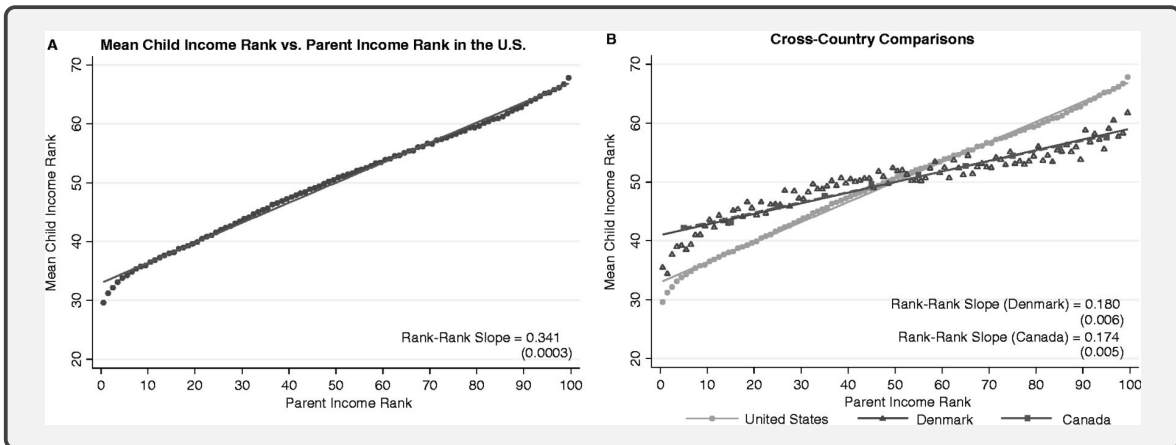


Fuente: Elaboración propia

Existe una menor proporción de entrevistados seleccionados en los grupos de ingreso más altos, lo cual es un reflejo de la distribución nacional, es decir, la población que percibe ingresos altos es menor, respecto a la que percibe ingresos bajos. Este mismo fenómeno ocurre en otros países, y es por ello que el promedio de rango percentil de los individuos con PPS en el rango percentil más alto siempre queda por debajo de 100.²⁴ En la Figura 5.8, se muestra el modelo de AIR para, EUA, Dinamarca y Canadá, tomado de Chetty y col. (2014).

²⁴ No hay individuos suficientes para que levanten el modelo AIR en la parte derecha.

Figura 5.8. Modelo de AIR para: EUA, Dinamarca y Canadá.



Fuente: Chetty y col. (2014)

Como se puede observar en la Figura 5.8, existe una movilidad promedio ascendente de los quintiles inferiores, mayor que la obtenida para México. De igual forma que en el país, los individuos actuales en los quintiles superiores son escasos. En estos países desarrollados observamos que la imagen de la recta del modelo AIR, tiene contenida a la imagen de recta del modelo AIR para México, lo cual tienen sentido en la comparación relativa.

5.5. Alcance de resultados

Los resultados se debe de interpretar con cautela, pues están determinados por todo el contexto ya descrito, es decir, la muestra pequeña de entrevistados de E17, y el error intrínseco de estimación tanto de los ingresos actuales, como los estimados con el método de MC2E2M.

Cabe destacar que las variables para la estimación de ingresos tanto actuales como de los PPS, también influyen en las posibles desviaciones que se podrían tener. Una posible solución a este problema sería utilizar más variables independientes (características del hogar, ubicación geográfica de los individuos, sector de la ocupación, etc.), pero esto disminuiría más la muestra utilizada. Por otro lado, se podrían explorar otras técnicas de ML para la estimación de ingresos actuales, de tal forma que la muestra no disminuya.

Si bien se obtienen resultados que pueden apoyarse de información adicional (densidades de ingresos y percepción de movilidad), existen algunos valores cuya interpretación

no es clara. En particular, se observa en la fila de Q_4 en la Tabla 5.3, que un individuo actual tiene casi la misma probabilidad de estar en alguno de los quintiles q_1, q_2, q_3, q_4 , si es que provienen de un PPS ubicado en Q_4 . Estos patrones pueden deberse también al tamaño de la muestra dentro de cada combinación de quintiles del PPS y del individuo actual.

Otro valor que no es claro de interpretar es la $P(q_4|Q_1) = 0.335$ de la región norte en la Tabla 5.5, lo cual indica que la probabilidad de estar en el quintil q_4 de un individuo actual, dado que su PPS se encontraba en Q_1 es 0.335, lo cual es un valor muy alto para la situación que describe, además es el valor más grande de la fila de Q_1 . Esto puede ser consecuencia de lo comentado al principio de esta sección y porque en la región norte no se tienen registros de individuos cuyo origen sea un PPS en el primer decil, entonces este comportamiento promedio no se ve disminuido.

Los resultados nacionales y regionales obtenidos parecen muy distantes de los reportados en la literatura,²⁵ el detalle está en que las comparaciones que se han establecido en lo general, pues difieren en metodología, bases de datos utilizadas y generación de nacimiento de individuos consideradas.

Los estudios de movilidad de Báez (2021) y Campos-Vázquez y col. (2020), toman individuos que nacieron aproximadamente entre 1970-1990, mientras que en este trabajo, se tienen individuos que en promedio nacieron entre 1980-1998 (esto es sin considerar todas las particularidades subyacentes de los métodos). En Campos-Vázquez y col. (2020) se hace referencia a una mayor movilidad en grupos más jóvenes, y precisamente bajo esa especificación es que los resultados aquí obtenidos son cercanos, como lo muestra la Figura 5.9.

²⁵ Algunos de estos valores se indican en la sección 5.1.

Figura 5.9. El cuadro en rojo resalta la estimación de la EII, para los grupos de nacimiento 1981-85 y 1986-90.

Table 3. Time-Variant and Cohort-Variant IGE Estimates					
<i>Dependent variable: log of adult children's earnings</i>					
	(1)	(2)	(3)	(4)	(5)
<i>log(SFE)</i>	–	–	–	–	0.468*** (0.033)
<i>x census year</i>					
1995	0.551*** (0.048)	0.770*** (0.058)	–	–	–
2000	0.514*** (0.061)	0.673*** (0.053)	–	–	-0.095 (0.080)
2010	0.322*** (0.037)	0.470*** (0.040)	–	–	-0.333*** (0.073)
2015	0.417*** (0.029)	0.474*** (0.033)	–	–	-0.345*** (0.085)
<i>x birth cohort</i>					
1971-75	–	–	0.525*** (0.034)	0.504*** (0.037)	–
1976-80	–	–	0.424*** (0.031)	0.390*** (0.031)	-0.024 (0.023)
1981-85	–	–	0.379*** (0.034)	0.326*** (0.051)	.003 (0.035)
1986-90	–	–	0.324*** (0.031)	0.290*** (0.053)	0.048 (0.041)
Cohort F.E.	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes
Region of birth F.E.	Yes	Yes	Yes	Yes	Yes
Age	No	Yes	No	Yes	Yes
Earnings profile	No	Yes	No	Yes	Yes
Observations	2,641,109	2,641,109	2,641,109	2,641,109	2,641,109
R ²	0.10	0.11	0.10	0.11	0.11

Fuente: Campos-Vázquez y col. (2020)

Conclusiones

En este trabajo se estudió la movilidad intergeneracional de México y sus regiones. Para ello se estimaron dos indicadores de movilidad, la elasticidad intergeneracional del ingreso y la asociación intergeneracional de rango.

Se tomaron datos de las encuestas ESRU - EMOVI 2017 y ENIGH 1994, 1996, 1998, 2016, 2018, y se construyeron datos de ingreso con apoyo de algoritmos de aprendizaje automático, con el fin de implementar el método de mínimos cuadrados en dos muestras en dos etapas, para la estimación de los indicadores de movilidad.

A nivel nacional y regional, tanto la estimación de la elasticidad intergeneracional del ingreso, como la de la asociación intergeneracional del rango, hacen referencia a una mayor movilidad ascendente por parte de los quintiles inferiores, pero esta movilidad en promedio está acotada en niveles bajos de ingreso.

Los resultados nacionales son cercanos a los reportados en Delajara y col. (2018), donde se indica que el 70 % de los Mexicanos que provienen del quintil más bajo no logran superar el segundo quintil. En los resultados encontrados se obtuvo que esto ocurre al 61 %.

De igual forma a Delajara y col. (2018), los resultados a nivel cualitativo apuntan a que la región sur es la que se encuentra en las peores condiciones de movilidad, y la región norte es la que se encuentra en mejores condiciones de movilidad.

Cabe mencionar que en la región sur hay una persistencia considerable en el quintil inferior, y una movilidad descendente de individuos cuyo PPS proviene del quintil superior,

pero este último resultado podría no ser contundente.

Lo expuesto anteriormente sugiere que aún hay áreas por explorar. Por un lado, los resultados sobre una sociedad relativamente móvil son consistentes con las colas izquierdas anchas de densidades de distribución y percepción de movilidad de los entrevistados. Sin embargo, la necesidad de estimar los ingresos, el tamaño reducido de la muestra, la consideración de movilidad tanto de hombres como de mujeres y una posible subrepresentación de individuos con ingresos altos, pueden estar sesgando (hacia bajo, de acuerdo a lo analizado) los resultados de movilidad económica intergeneracional.

Para futuras investigaciones, se puede considerar una construcción más robusta sobre la estimación de los ingresos actuales de los entrevistados. Se pueden contrastar las relaciones de ingresos familiar actual contra la familia de origen, pero de igual forma que en el trabajo actual, habría un reto en la estimación de ingresos. Esto resalta la necesidad de generar más y mejores datos que permitan dar seguimiento a los resultados económicos de las generaciones.

Bibliografía

- Báez, N. A. D. (2021). *Intergenerational Earnings Mobility in Mexico* (DoQSS Working Papers N.º 21-10). Quantitative Social Science - UCL Social Research Institute, University College London. <https://ideas.repec.org/p/qss/dqsswp/2110.html>
- Banco de México. (2011-2021). *Reportes sobre las economías regionales, estados, Banco de México*. Consultado el 28 de febrero de 2021, desde <https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-las-economias-regionales/reportes-economias-regionales.html>
- Blanden, J. (2019). Intergenerational income persistence. *IZA World of Labor*. Consultado el 6 de enero de 2021, desde <https://wol.iza.org/articles/intergenerational-income-persistence/long>
- Campos-Vázquez, R. M., Delgado Barrera, V. H. & Vélez-Grajales, R. (2020). *Intergenerational Economic Mobility in Mexico* (inf. téc.). Consultado el 7 de enero de 2021, desde <https://ceey.org.mx/intergenerational-economic-mobility-in-mexico/>
- Chetty, R., Hendren, N., Kline, P. & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, 129(4), 1553-1623. <https://doi.org/10.1093/qje/qju022>
- Delajara, M., De la Torre, R., Díaz-Infante, E. & Vélez, R. (2018). *El México del 2018: Movilidad Social Para el Bienestar*. Centro de Estudios Espinosa Yglesias. Consultado el 28 de enero de 2021, desde <https://ceey.org.mx/el-mexico-del-2018/>
- Delajara, M. & Graña, D. (2018). Intergenerational Social Mobility in Mexico and its Regions: Results from Rank-Rank Regressions. *Centro de Estudios Espinosa Yglesias*. Consultado el 30 de abril de 2021, desde <https://ceey.org.mx/intergenerational-social-mobility-in-mexico-and-its-regions/>

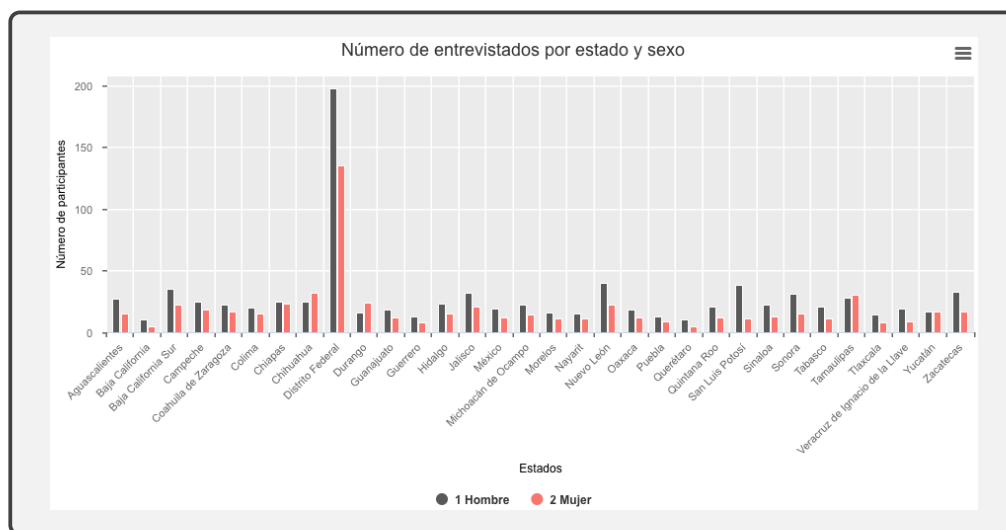
- Mocetti, S. (2007). Intergenerational Earnings Mobility in Italy. *The B.E. Journal of Economic Analysis & Policy*, 7(2), 1-25. <https://ideas.repec.org/a/bpj/bejeap/v7y2007i2n5.html>
- Monroy-Gómez-Franco, L. & Corak, M. (2020). *A Land of Unequal Chances: Social Mobility Across Mexican Regions* (Documento de trabajo N.º 10/2020). Centro de Estudios Espinosa Yglesias. <https://ceey.org.mx/a-land-of-unequal-chances-social-mobility-across-mexican-regions/>

Bases de datos

- CEEY. (2018). ESRU – EMOVI. Consultado el 7 de enero de 2021, desde <https://ceey.org.mx/contenido/que-hacemos/emovi/>
- INEGI-ENIGH. (1994). Consultado el 7 de enero de 2021, desde <https://www.inegi.org.mx/programas/enigh/tradicional/1994/>
- INEGI-ENIGH. (1996). Consultado el 7 de enero de 2021, desde <https://www.inegi.org.mx/programas/enigh/tradicional/1996/>
- INEGI-ENIGH. (1998). Consultado el 7 de enero de 2021, desde <https://www.inegi.org.mx/programas/enigh/tradicional/1998/>
- INEGI-ENIGH. (2016). Consultado el 7 de enero de 2021, desde <https://www.inegi.org.mx/programas/enigh/nc/2016/>
- INEGI-ENIGH. (2018). Consultado el 7 de enero de 2021, desde <https://www.inegi.org.mx/programas/enigh/nc/2018/>

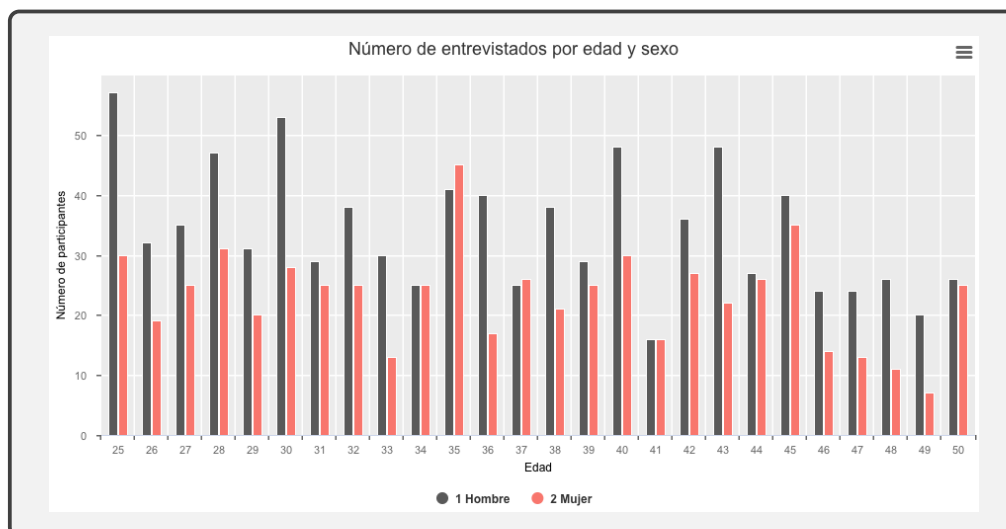
Estadísticas ESRU - EMOVI

Figura A.1. Número de entrevistados seleccionados por estado y sexo, criterio A.



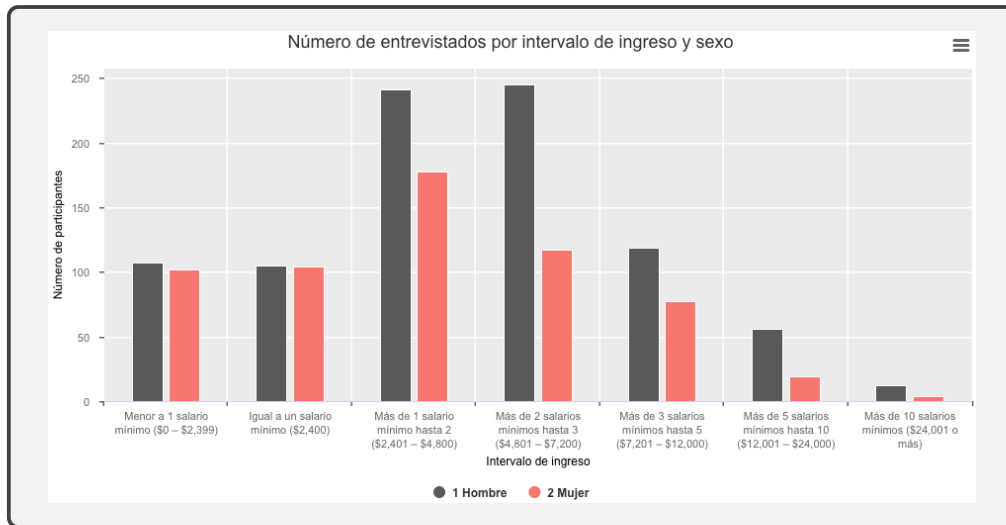
Fuente: Elaboración propia

Figura A.2. Número de entrevistados seleccionados por edad y sexo, criterio A.



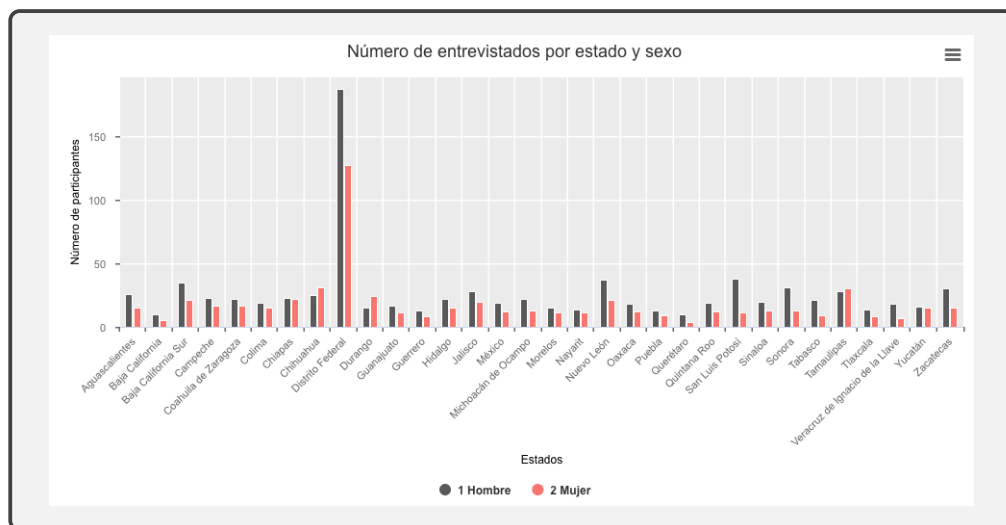
Fuente: Elaboración propia

Figura A.3. Número de entrevistados seleccionados por intervalo de ingreso y sexo, criterio A.



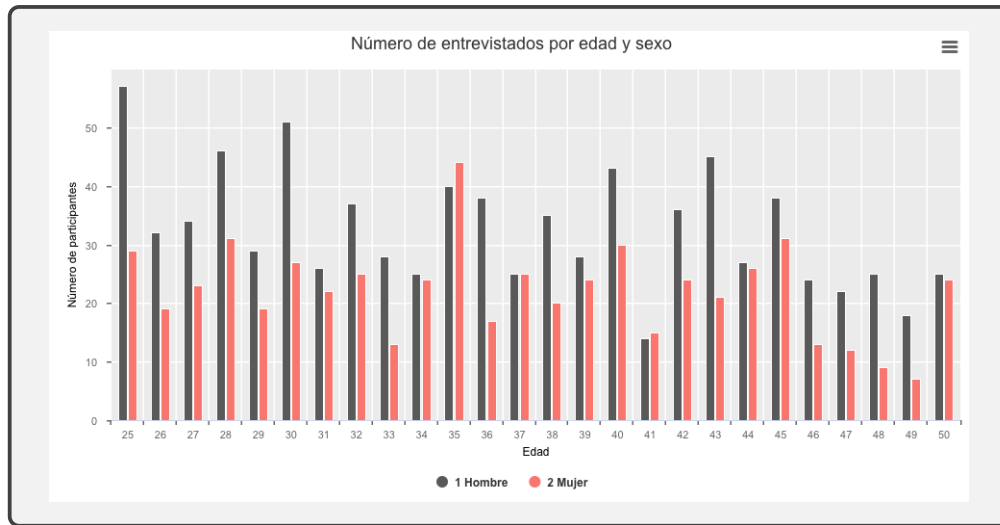
Fuente: Elaboración propia

Figura A.4. Número de entrevistados seleccionados por estado y sexo, criterio B.



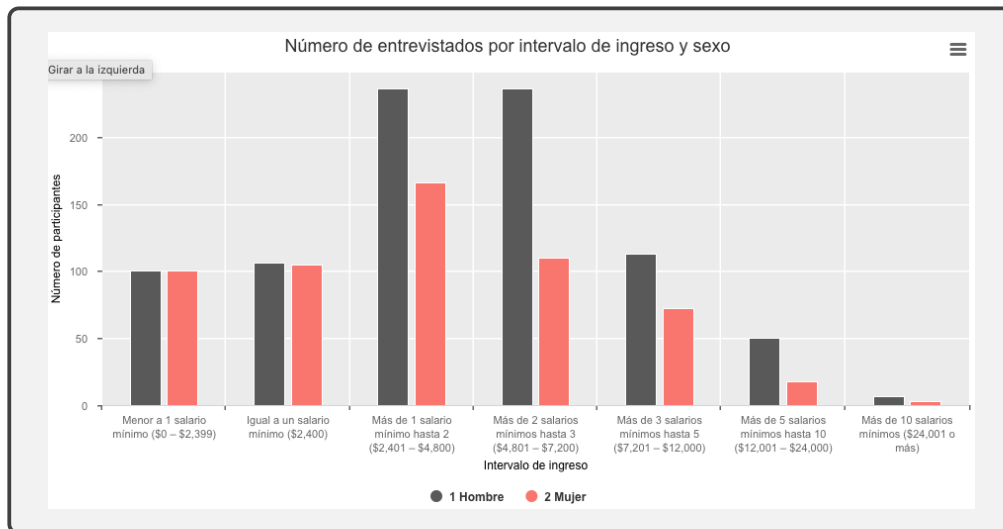
Fuente: Elaboración propia

Figura A.5. Número de entrevistados seleccionados por edad y sexo, criterio B.



Fuente: Elaboración propia

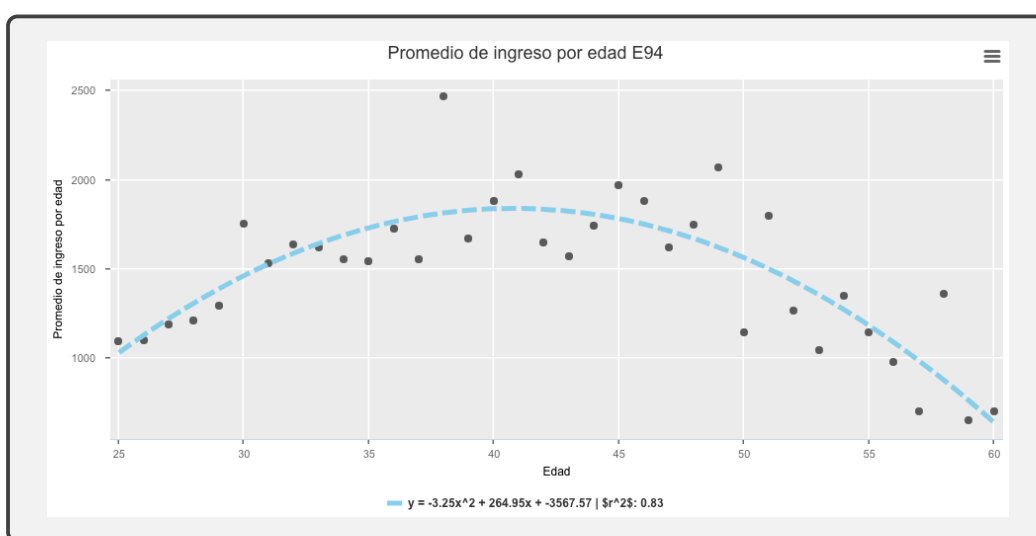
Figura A.6. Número de entrevistados seleccionados por intervalo de ingreso y sexo, criterio B.



Fuente: Elaboración propia

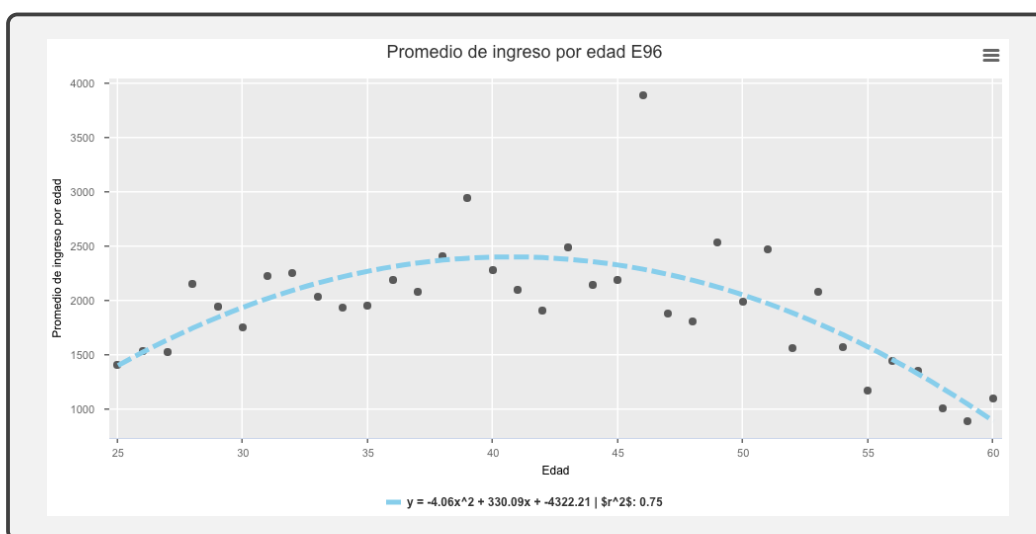
Estadísticas ENIGH

Figura B.1. Promedio de ingresos por edad para E94.



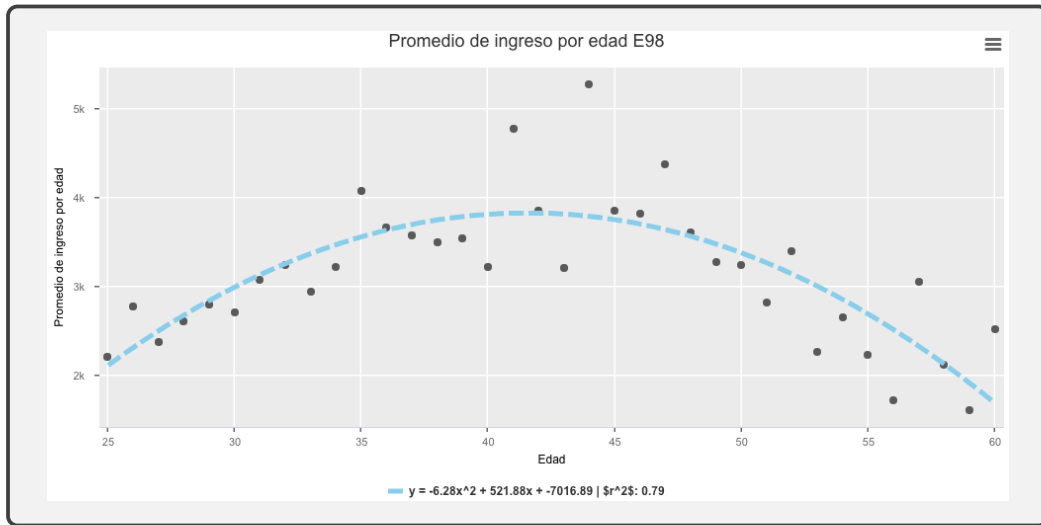
Fuente: Elaboración propia

Figura B.2. Promedio de ingresos por edad para E96.



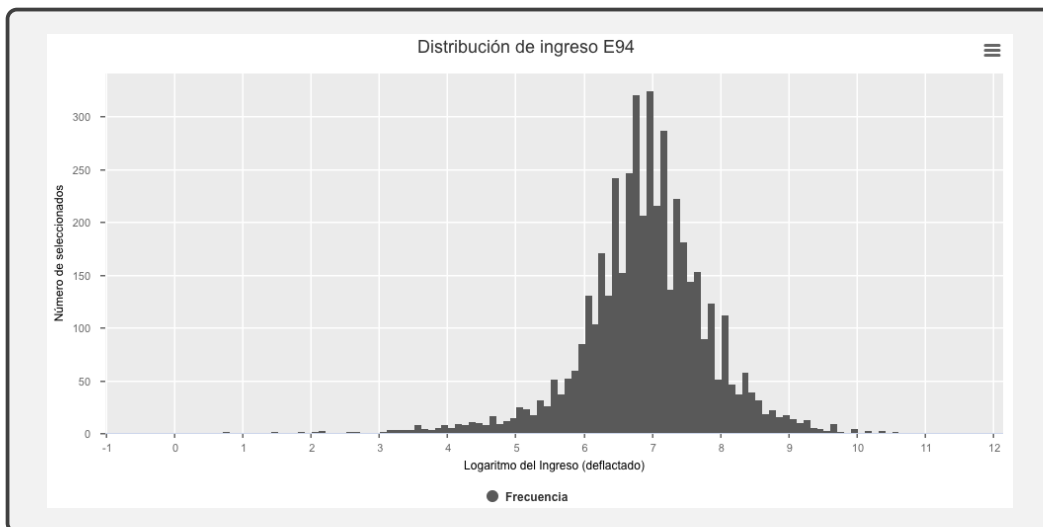
Fuente: Elaboración propia

Figura B.3. Promedio de ingresos por edad para E98.



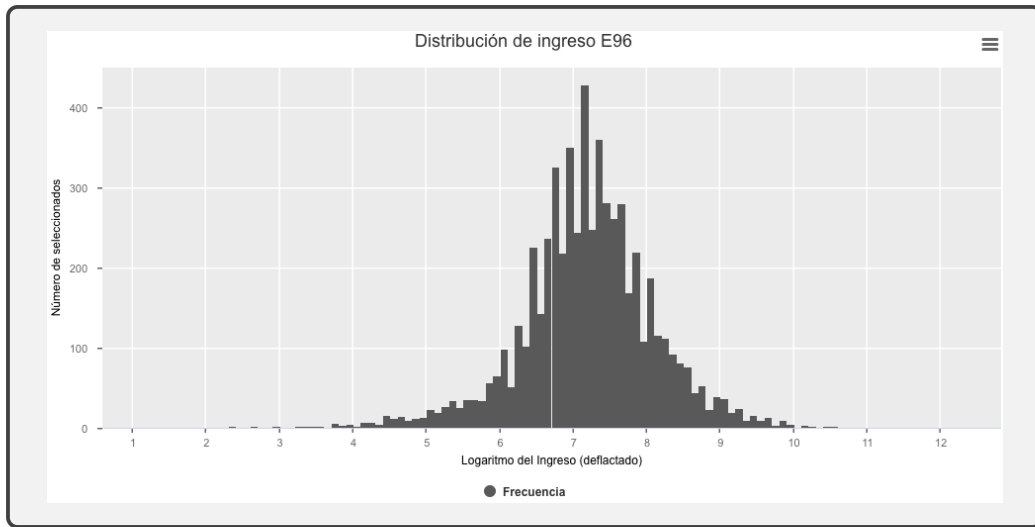
Fuente: Elaboración propia

Figura B.4. Histograma de ingreso E94 (logaritmo).



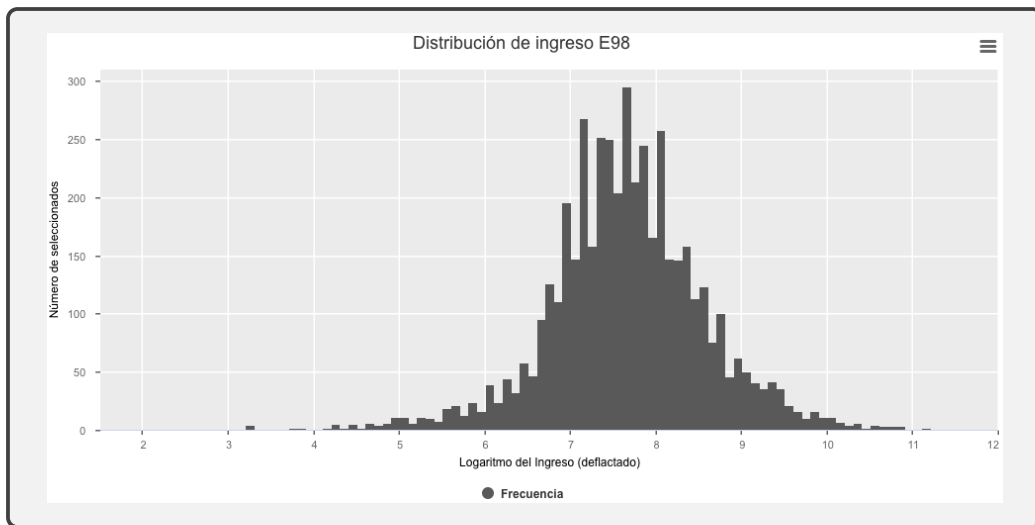
Fuente: Elaboración propia

Figura B.5. *Histograma de ingreso E96 (logaritmo).*



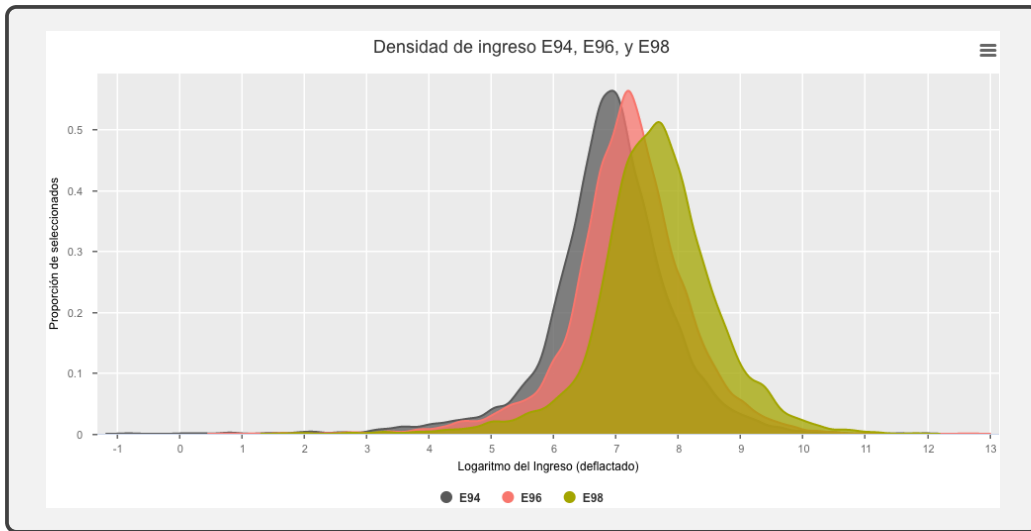
Fuente: Elaboración propia

Figura B.6. *Histograma de ingreso E98 (logaritmo).*



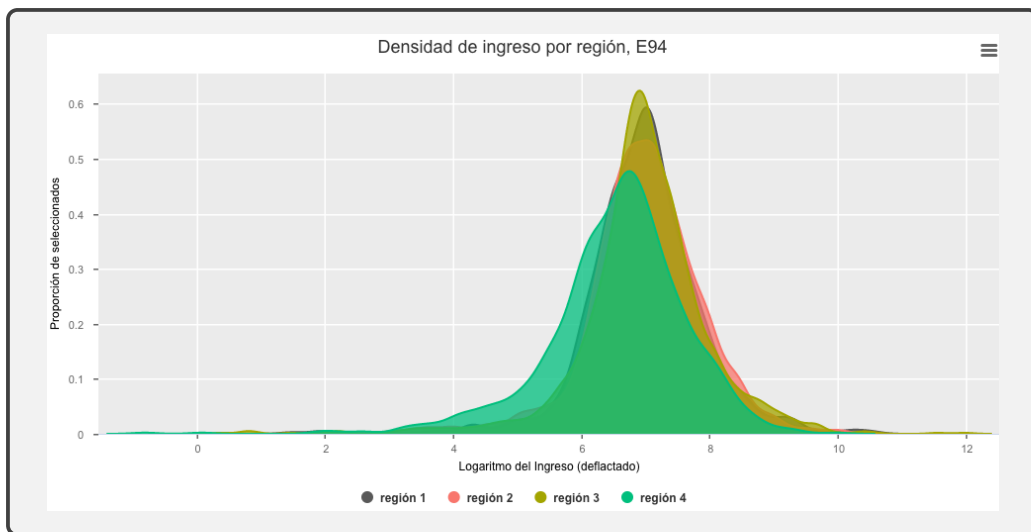
Fuente: Elaboración propia

Figura B.7. Densidad de ingreso empleando las muestras E94, E96 y E98.



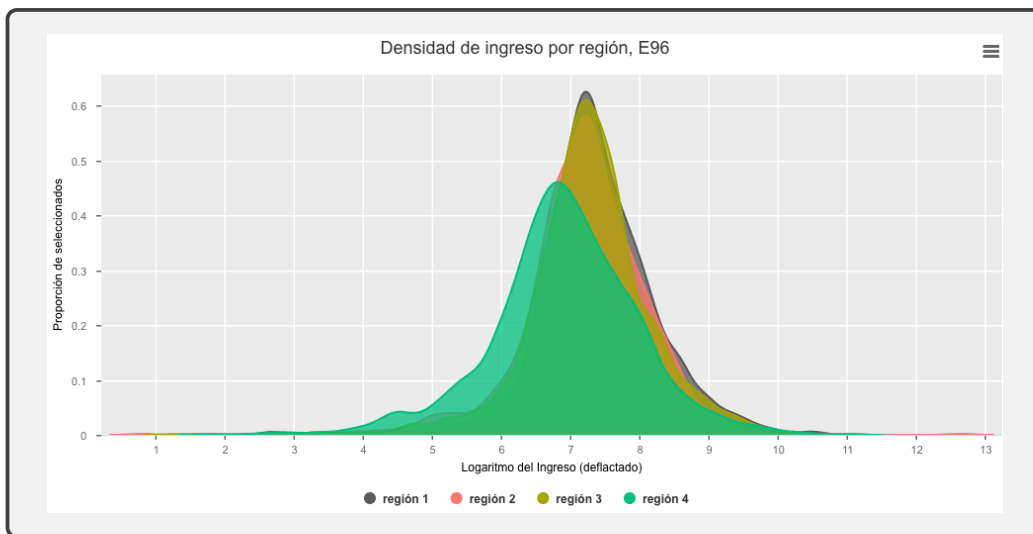
Fuente: Elaboración propia

Figura B.8. Densidad de ingreso por región para E94.



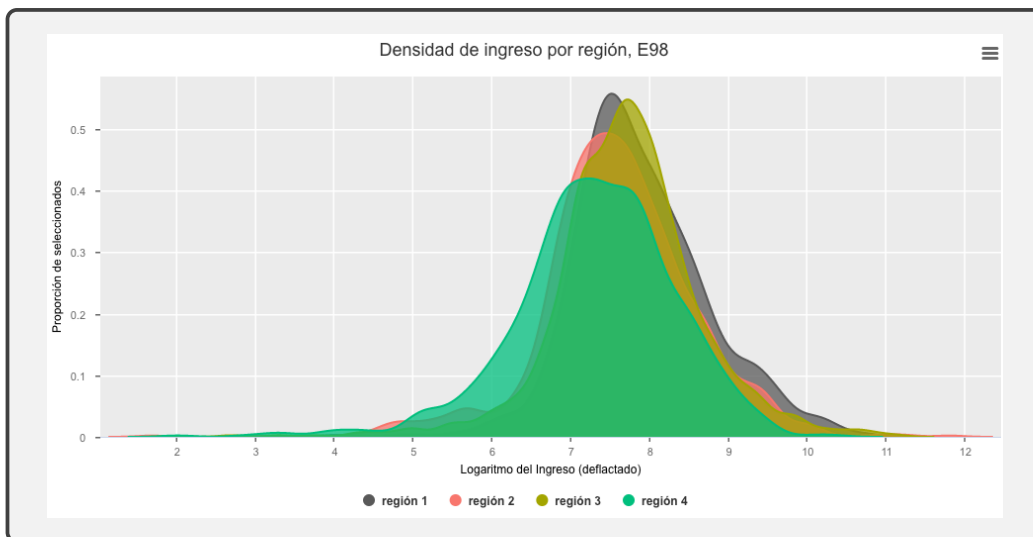
Fuente: Elaboración propia

Figura B.9. Densidad de ingreso por región para E96.



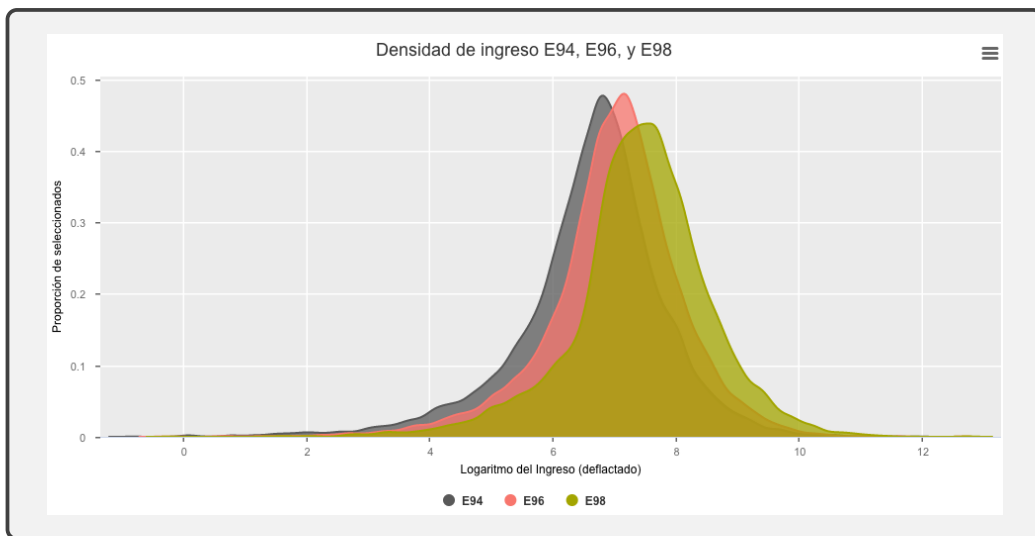
Fuente: Elaboración propia

Figura B.10. Densidad de ingreso por región para E98.



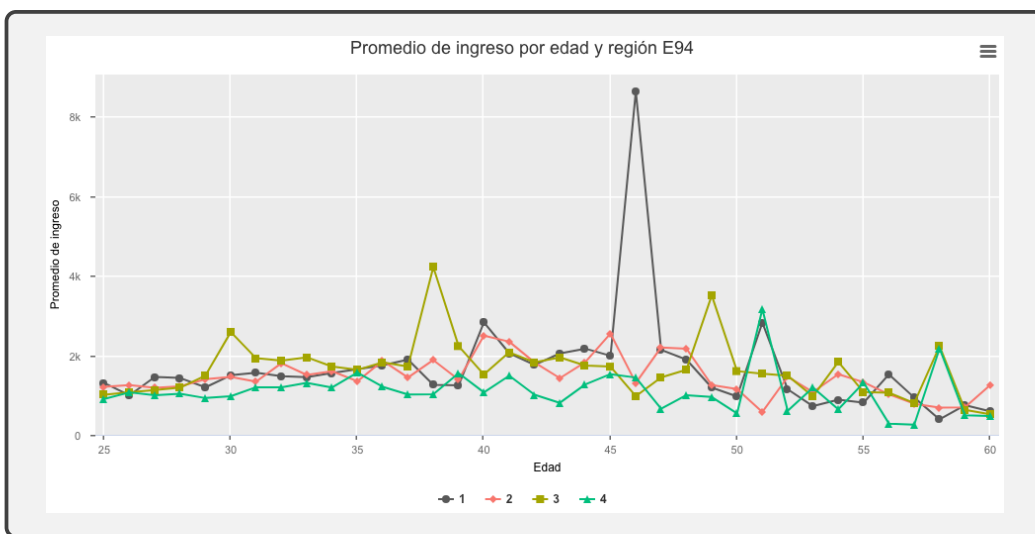
Fuente: Elaboración propia

Figura B.11. Densidad de ingreso de la población para E94, E96 y E98.



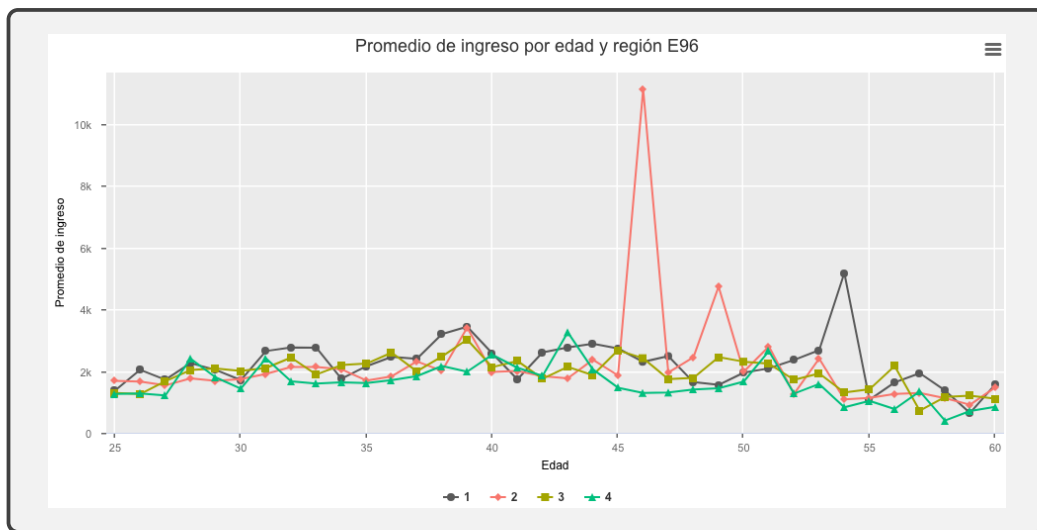
Fuente: Elaboración propia

Figura B.12. Promedio de ingreso por edad y región para E94.



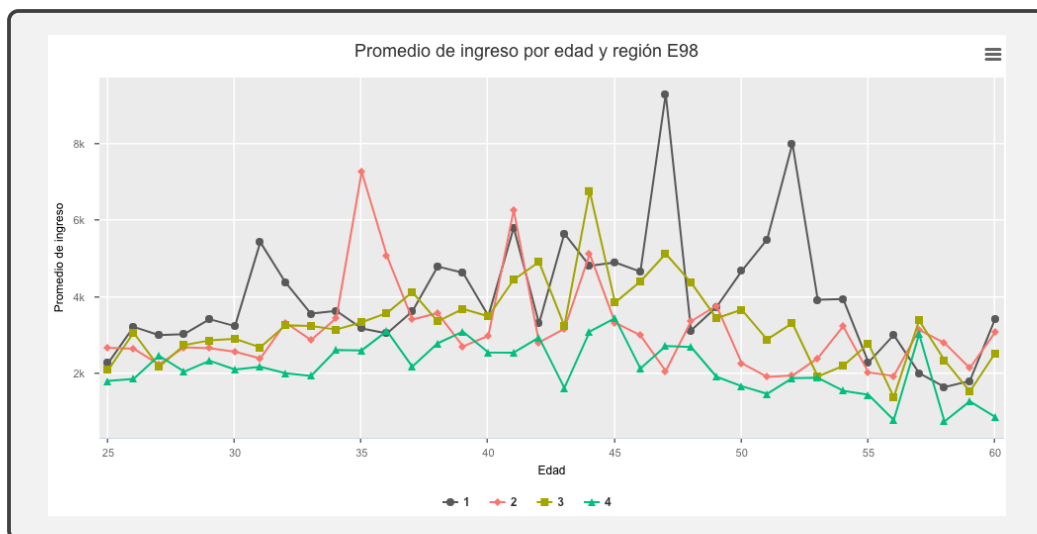
Fuente: Elaboración propia

Figura B.13. Promedio de ingreso por edad y región para E96.



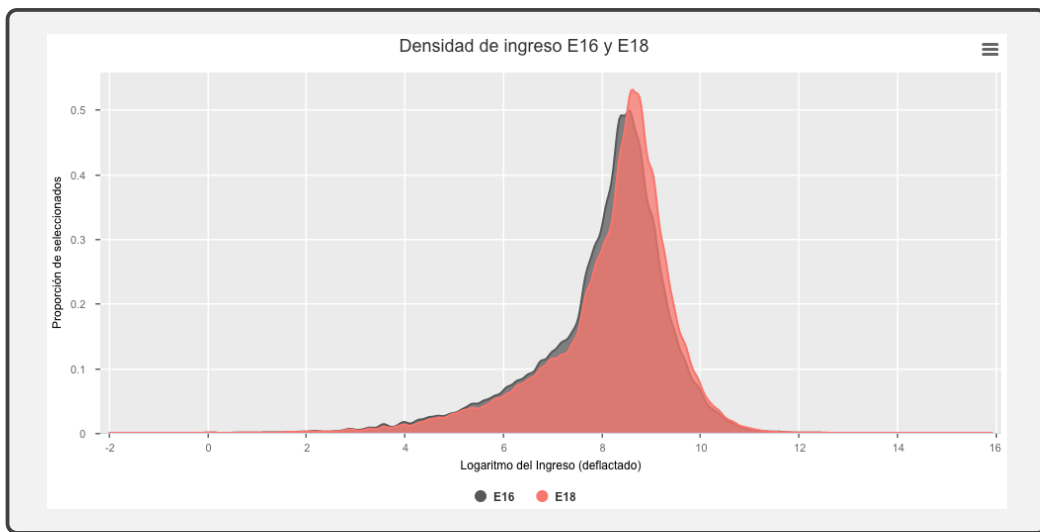
Fuente: Elaboración propia

Figura B.14. Promedio de ingreso por edad y región para E98.



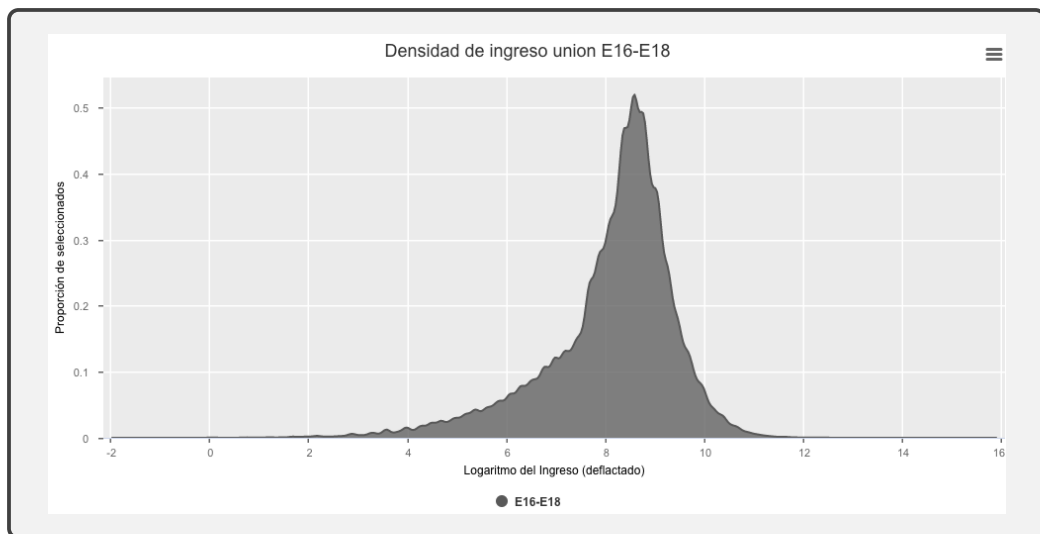
Fuente: Elaboración propia

Figura B.15. Densidad de ingreso de la población para E16 y E18.



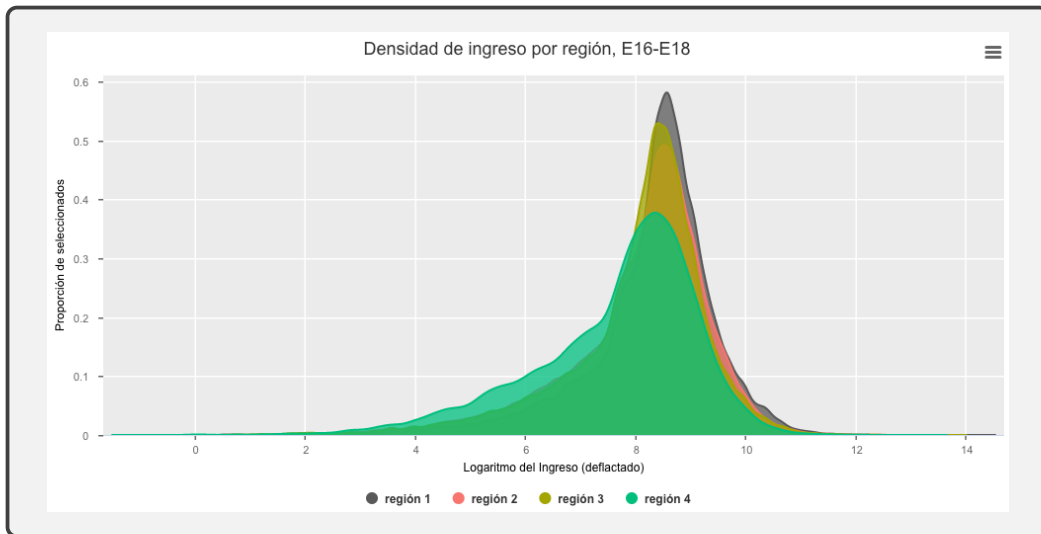
Fuente: Elaboración propia

Figura B.16. Densidad de ingreso de la población para la unión de E16 y E18.



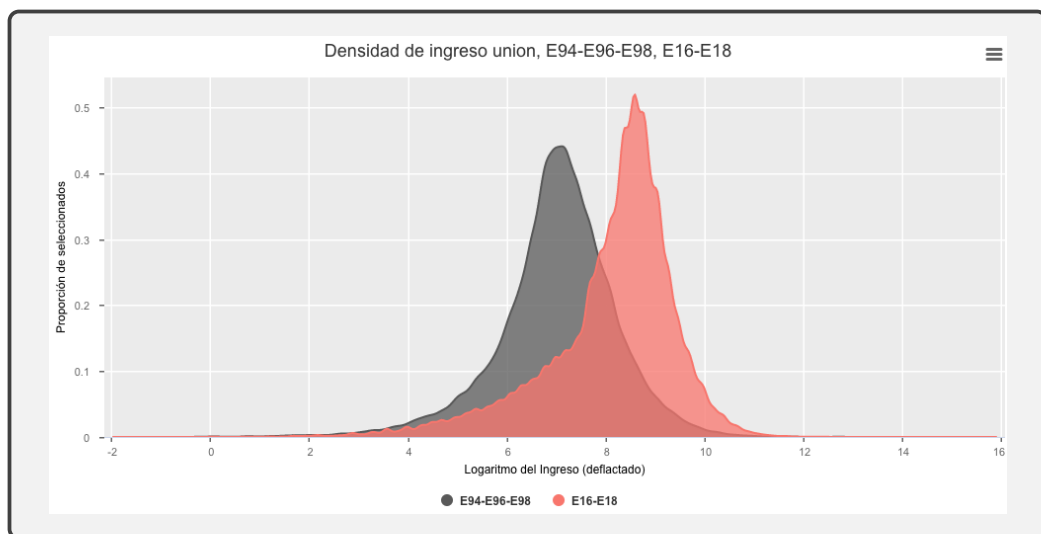
Fuente: Elaboración propia

Figura B.17. Densidad de ingreso de la población por región para la unión de E16 y E18.



Fuente: Elaboración propia

Figura B.18. Densidad de ingreso de la población E94-E96-E98, E16-E18.



Fuente: Elaboración propia

Clasificaciones

Tabla C.1. Codificación de ocupaciones, Clasificación Mexicana de Ocupaciones (CMO)

Ocupación	Código
Profesionistas.	11
Técnicos.	12
Trabajadores en la educación.	13
Trabajadores del arte, espectáculos y deportes.	14
Funcionarios y directivos de los sectores público, privado y social.	21
Trabajadores en actividades agrícolas, ganaderas, silvícolas, y de caza y pesca.	41
Jefes, supervisores y otros trabajadores de control en la fabricación artesanal.	51
Artesanos y trabajadores fabriles en la industria de la transformación, etc.	52
Operadores de maquinaria fija de movimiento continuo y equipos en el proceso de producción.	53
Ayudantes, peones y similares en el proceso de la fabricación artesanal e industrial	54
Conductores y ayudantes de conductores de maquinaria móvil y medios de transporte.	55
Jefes de departamento, coordinadores y supervisores en actividades administrativas y de servicios.	61
Trabajadores de apoyo en actividades administrativas.	62
Comerciantes, empleados de comercio y agentes de ventas.	71
Vendedores ambulantes y trabajadores ambulantes en servicios.	72
Trabajadores en servicios personales en establecimientos.	81
Trabajadores en servicios domésticos.	82
Trabajadores en servicios de protección y vigilancia y fuerzas armadas.	83
Otros trabajadores con ocupaciones no clasificados anteriormente, insuficientemente especificado.	99

Fuente: Elaboración propia

Tabla C.2. Codificación de años de educación

Años de educación	Código
0 a 5	C_1
6 a 10	C_2
11 y 12	C_3
13 y 14	C_4
15 o más	C_5

Fuente: Elaboración propia

Tabla C.3. Codificación de regiones

Región	Código
Norte (región 1)	1
Centro-Norte (región 2)	2
Centro (región 3)	3
Sur (región)	4

Fuente: Elaboración propia

Estimaciones

Tabla D.1. *Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango nacionales*

Estimación	ENIGH			
	1994	1996	1998	Promedio
<i>Panel A: Criterio A</i>				
$\hat{\beta}$	0.225*** (0.038)	0.249*** (0.039)	0.261*** (0.042)	0.247*** (0.038)
$\hat{\rho}$	0.2120*** (0.029)	0.223*** (0.029)	0.241*** (0.031)	0.226*** (0.029)
$\hat{\alpha}$	45.888*** (1.637)	45.735*** (1.642)	44.615*** (1.709)	45.363*** (1.598)
<i>Panel B: Criterio B</i>				
$\hat{\beta}$	0.341*** (0.061)	0.362*** (0.057)	0.368*** (0.067)	0.359*** (0.064)
$\hat{\rho}$	0.248*** (0.033)	0.254*** (0.031)	0.263*** (0.038)	0.262*** (0.034)
$\hat{\alpha}$	29.549*** (1.797)	29.682*** (1.671)	28.886*** (2.002)	28.937*** (1.873)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Fuente: *Elaboración propia*

Tabla D.2. Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango (criterio A) regionales.

<i>Región de México</i>				
<i>Panel A: 1994</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\beta}$	0.238** (0.121)	0.259*** (0.081)	0.188*** (0.060)	0.341*** (0.077)
$\hat{\rho}$	0.199*** (0.073)	0.234*** (0.059)	0.181*** (0.051)	0.335*** (0.060)
$\hat{\alpha}$	52.142*** (4.278)	44.324*** (3.430)	44.976*** (3.154)	37.741*** (2.692)
<i>Panel B: 1996</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\beta}$	0.253** (0.116)	0.282*** (0.083)	0.228*** (0.066)	0.313*** (0.081)
$\hat{\rho}$	0.231*** (0.078)	0.260*** (0.058)	0.209*** (0.050)	0.297*** (0.068)
$\hat{\alpha}$	49.596*** (4.943)	43.728*** (3.184)	43.792*** (2.706)	39.592*** (3.162)
<i>Panel C: 1998</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\rho}$	0.223** (0.111)	0.290*** (0.086)	0.235 (0.072)	0.356*** (0.084)
$\hat{\rho}$	0.196** (0.089)	0.274*** (0.065)	0.224*** (0.049)	0.328*** (0.063)
$\hat{\alpha}$	51.321*** (5.832)	43.386*** (3.378)	42.863*** (2.851)	38.382*** (2.579)
<i>Panel D: Promedio</i>	Norte	Centro-Norte	Centro	Sur
Estimación				
$\hat{\rho}$	0.254** (0.113)	0.295*** (0.076)	0.216*** (0.068)	0.343*** (0.080)
$\hat{\rho}$	0.216*** (0.073)	0.271*** (0.057)	0.208*** (0.048)	0.328*** (0.064)
$\hat{\alpha}$	50.513*** (4.642)	42.815*** (3.193)	43.817*** (2.835)	38.431*** (2.968)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Fuente: Elaboración propia

Tabla D.3. Estimación de la elasticidad intergeneracional del ingreso y de los parámetros de la asociación intergeneracional del rango (criterio B) regionales.

<i>Región de México</i>				
Panel A: 1994				
Estimación	Norte	Centro-Norte	Centro	Sur
$\hat{\beta}$	0.282 (0.178)	0.314** (0.133)	0.304*** (0.098)	0.220 (0.154)
$\hat{\rho}$	0.158 (0.108)	0.236*** (0.078)	0.253*** (0.060)	0.178** (0.070)
$\hat{\alpha}$	40.501*** (6.268)	28.646*** (4.131)	30.195*** (3.554)	23.430*** (3.169)
Panel B: 1996				
Estimación	Norte	Centro-Norte	Centro	Sur
$\hat{\beta}$	0.384** (0.169)	0.394*** (0.148)	0.317*** (0.100)	0.240* (0.139)
$\hat{\rho}$	0.228* (0.120)	0.284*** (0.075)	0.252*** (0.057)	0.175** (0.075)
$\hat{\alpha}$	35.576*** (7.255)	26.696*** (3.754)	31.076*** (3.505)	23.913*** (2.741)
Panel C: 1998				
Estimación	Norte	Centro-Norte	Centro	Sur
$\hat{\beta}$	0.295 (0.189)	0.342** (0.150)	0.324*** (0.108)	0.244 (0.161)
$\hat{\rho}$	0.207 (0.129)	0.245*** (0.077)	0.269*** (0.066)	0.179** (0.080)
$\hat{\alpha}$	36.589*** (8.514)	28.938*** (3.790)	29.962*** (3.691)	23.668*** (3.501)
Panel D: Promedio				
Estimación	Norte	Centro-Norte	Centro	Sur
$\hat{\beta}$	0.325** (0.157)	0.385*** (0.128)	0.310*** (0.114)	0.236 (0.154)
$\hat{\rho}$	0.198* (0.113)	0.274*** (0.080)	0.255*** (0.067)	0.164** (0.076)
$\hat{\alpha}$	37.525*** (6.708)	27.247*** (4.391)	30.472*** (3.942)	23.971*** (3.244)

Nota: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

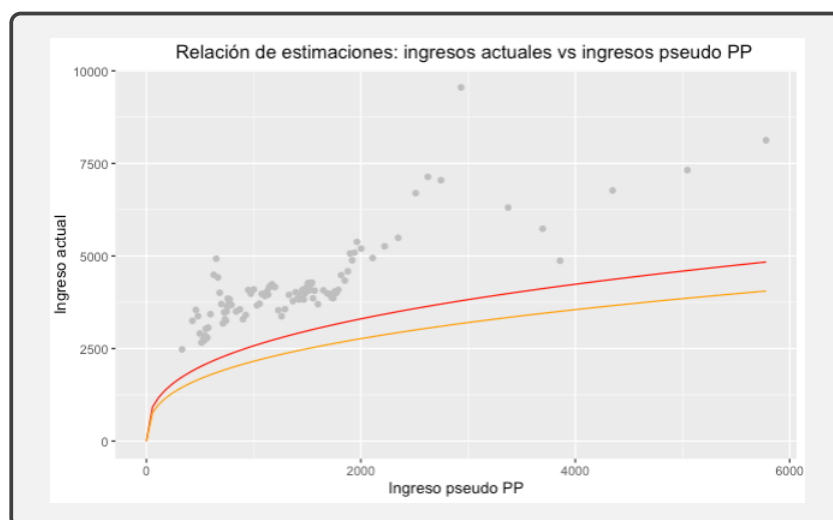
Fuente: Elaboración propia

Relación de ingresos estimados

Para la construcción de la gráfica de ingresos estimados de la generación actual contra los ingresos estimados de los PP se sigue a (Chetty y col., 2014), se divide en centiles ambos ingresos y se promedian los salarios que caen dentro de cada centil.

En la Figura E.1, se observa que la relación entre los ingresos es no lineal, motivo por el cual el modelo logarítmico es la propuesta en MC2E2M. La curva en rojo es la transformación exponencial de la segunda etapa de MC2E2M sin considerar el control por sexo, y la curva en naranja es considerando el control por sexo. Se observa que en promedio ser mujer tienen una menor proyección de ingresos respecto de su pseudo PP

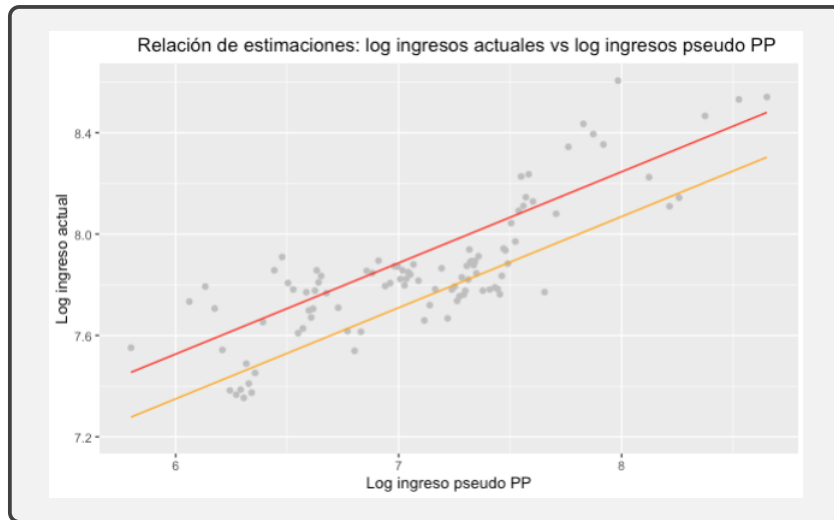
Figura E.1. *Relación entre ingreso actual y PPS.*



Fuente: Elaboración propia

Si se elabora la transformación logarítmica el resultado es el siguiente

Figura E.2. *Relación entre el logaritmo de ingreso actual y PPS.*

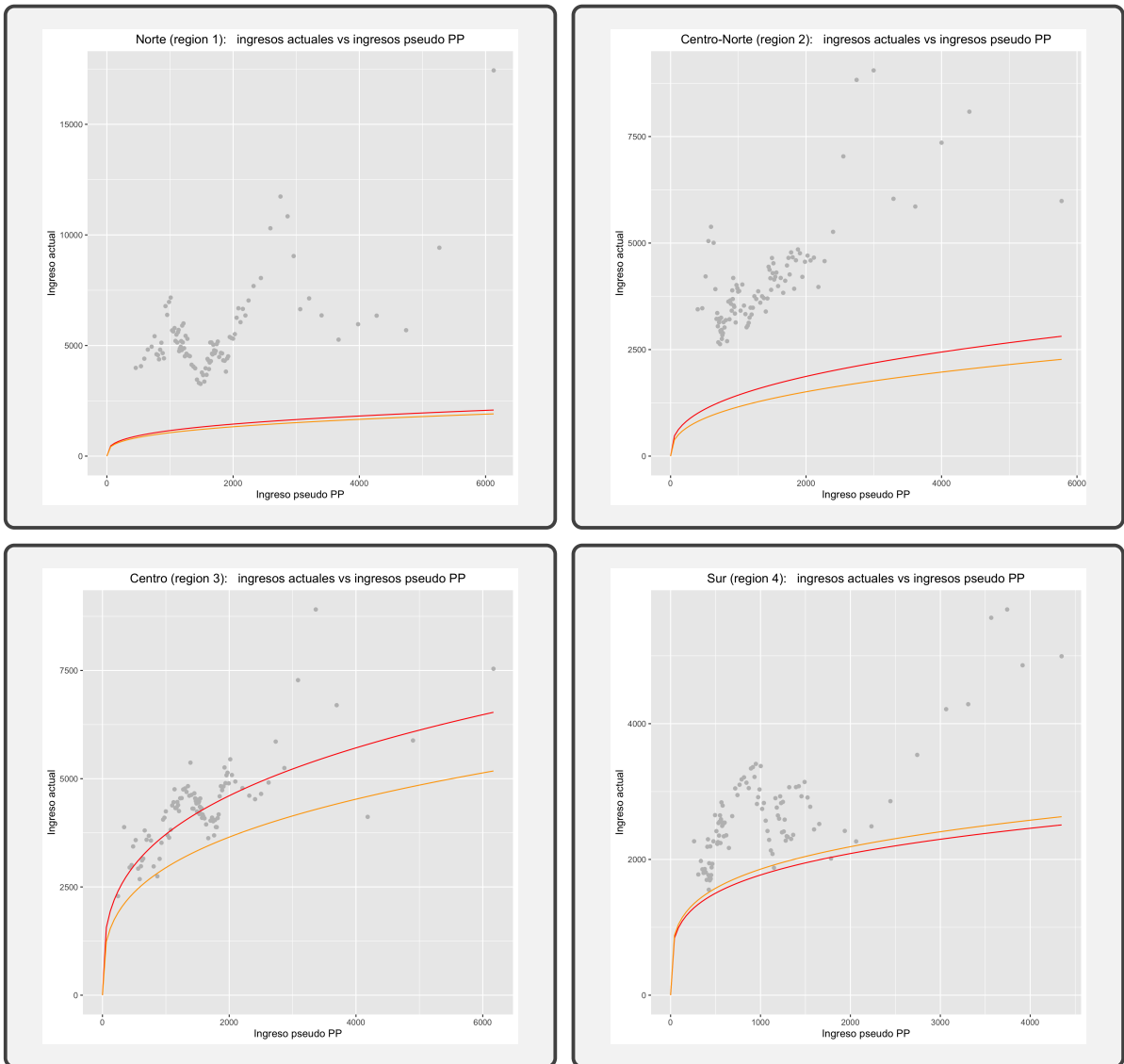


Fuente: Elaboración propia

Como se puede observar la estimación es correcta bajo los datos disponibles. Entonces, bajo estas circunstancias, pensar en la existencia de movilidad parece una consecuencia de las observaciones.

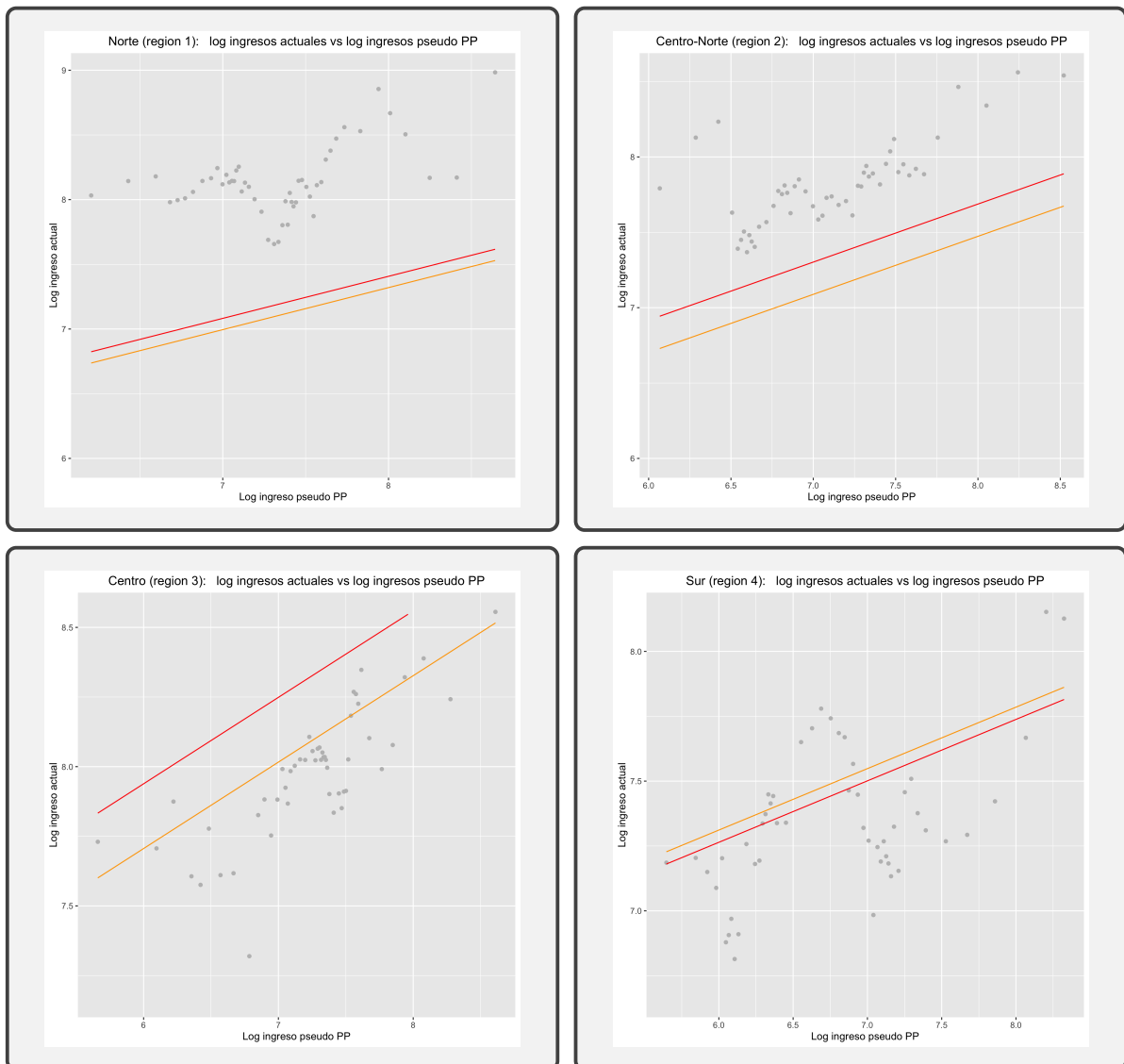
El análisis similar para las regiones se muestra en las Figuras [E.3](#) y [E.4](#)

Figura E.3. *Relación de ingresos estimados.*



Fuente: Elaboración propia

Figura E.4. Relación del logaritmo de ingresos estimados por región.



Fuente: Elaboración propia