

Impacto de la Educación, Género y Composición Familiar en el Ingreso de los Hogares Mexicanos: Un Enfoque Bietápico

Camacho Sánchez Ian Alejandro Garrido Cruz Azul Estefanía
Hernández Elizarrarás Edgar Antonio

30 de Mayo, 2025

Abstract Understanding the determinants of household income is a central issue in the study of economic inequality, as income is shaped by individual characteristics, household attributes, and the broader socioeconomic context. Despite an extensive body of literature, significant gaps remain—particularly regarding how these determinants vary across the income distribution and the potential endogeneity of key explanatory variables. This study employs a two-stage model using microdata from Mexican households to examine how factors such as the household head’s education and gender, family composition, and income decile influence both current income and the likelihood of participating in the informal labor market.

Keywords: Household income, linear models, socioeconomic inequality, education, income deciles, Mexican households, formal work.

JEL Classification: *D31, I24, C31, R20*

Resumen La comprensión de los determinantes del ingreso de los hogares constituye un eje central en el análisis de la desigualdad económica, dado que dicho ingreso está influenciado por características individuales, atributos del hogar y el entorno socioeconómico en el que se insertan. A pesar de la extensa literatura existente, persisten vacíos importantes, particularmente en relación con la forma en que estos determinantes varían a lo largo de la distribución del ingreso y respecto a la potencial endogeneidad de algunas variables clave. Este estudio emplea un modelo de dos etapas con microdatos de hogares mexicanos para examinar cómo factores como el nivel educativo y el género del jefe del hogar, la composición familiar y el decil de ingreso inciden tanto en el ingreso corriente como en la probabilidad de participar en el mercado laboral informal.

Palabras Clave: Ingreso de los hogares, modelos lineales, desigualdad socioeconómica, educación, deciles de ingreso, hogares mexicanos, trabajo formal.

Código JEL: *D31, I24, C31, R20*

Introducción

La comprensión de los determinantes del ingreso de los hogares ha sido un tema central en los estudios de desigualdad económica, donde se reconoce que el ingreso depende tanto de características individuales como de factores propios del hogar y del contexto socioeconómico en que estos se desenvuelven (Heckman y Landerso, 2022; Muluneh y Addisu, 2024 ; Van Vu, 2020). Diversas investigaciones han examinado el papel de la educación en el ingreso y la calidad de vida (Albert y Ángeles, 2005; Heckman y Landerso, 2022; Van Vu, 2020). Por ejemplo, Albert y Ángeles (2005) encontraron que la educación se relaciona positivamente con la satisfacción laboral y los ingresos en los países del sur de la Unión Europea, destacando cómo la educación mejora la calidad laboral y los ingresos, especialmente en contextos de bajos salarios. Asimismo, Heckman et al. (2014) destacan que el nivel educativo tiene efectos significativos en diversas esferas, incluidas la salud y el mercado laboral, subrayando cómo factores cognitivos y socioemocionales también influyen en los resultados educativos y laborales. Estos hallazgos resaltan la relevancia de analizar cómo las características de los individuos y del hogar influyen en el ingreso, en particular en contextos socioeconómicos variados.

No obstante, persisten importantes vacíos en la literatura. La mayoría de los estudios tienden a abordar los efectos de manera agregada, sin discutir suficientemente la presencia de endogeneidad. En países en desarrollo, la investigación de Van Vu (2020) sobre hogares rurales en Vietnam muestra que, si bien la educación incrementa los ingresos, los beneficios son mayores en los hogares con ingresos más altos, lo que agrava la desigualdad en áreas rurales. Esta variabilidad en los efectos de la educación sobre el ingreso sugiere la necesidad de profundizar en estudios que segmenten a los hogares según su nivel de ingreso y por tanto los ingresos de los hogares formales e informales. En este sentido, Fusco e Islam (2017) identifican que el tamaño del hogar y la presencia de hijos aumentan la probabilidad de pobreza, especialmente en contextos donde los ingresos son bajos, acentuando aún más la relevancia de considerar las características del hogar en el análisis de ingreso. Por otro lado, Muluneh y Addisu (2024) identificaron que la diversificación de ingresos en hogares urbanos de Etiopía se relaciona positivamente con la seguridad alimentaria, pero observan que factores como la educación y el acceso a recursos financieros son determinantes críticos, revelando que el contexto influye fuertemente en la capacidad de los hogares para mejorar su situación económica.

Además, investigaciones en países desarrollados como Dinamarca y Estados Unidos, comparadas por Heckman y Landerso (2022), sugieren que las políticas universales de bienestar no siempre eliminan las desigualdades intergeneracionales, ya que los hogares con mayor capacidad económica y recursos tienden a aprovechar mejor estas políticas, perpetuando así ciertas ventajas. Estas evidencias ponen de manifiesto que los factores económicos y sociales que influyen en el ingreso de los hogares requieren un análisis detallado que considere tanto las características individuales como las diferencias entre niveles de ingreso.

Este estudio, por tanto, tiene como objetivo analizar, mediante una serie de modelos específicos, cómo el ingreso corriente en los hogares mexicanos es influenciado por características individuales de los jefes del hogar y factores propios del hogar, así como la presencia

de autoselección cuando llegamos a analizar la formalidad de los hogares. Este enfoque permitirá descomponer los efectos de variables como la educación, el género del jefe de hogar y la composición familiar, y observar cómo estos afectan de manera diferenciada a los hogares formales, atendiendo a las especificidades de cada decil. Finalmente, al implementar un modelo Heckman, este estudio contribuye al campo al proporcionar una visión desagregada y detallada de los determinantes del ingreso, informando sobre cómo las políticas socioeconómicas pueden adaptarse para abordar desigualdades estructurales específicas en cuanto a la formalidad laboral en un primer momento.

El presente estudio se organiza en varias secciones que permiten comprender en profundidad los aspectos planteados. La primera sección ofrece una revisión de la literatura existente, analizando cómo otros estudios han abordado este tema y resaltando las limitaciones que diferencian sus enfoques del nuestro. En la segunda sección, describimos las herramientas utilizadas. La tercera sección continúa incluyendo un análisis de los datos empleados. La cuarta sección se dedica a exponer los resultados principales de la estimación econométrica. En la quinta sección se discuten estos hallazgos y concluimos.

Primera Sección: Marco Teórico

La teoría económica señala que los individuos tienen el objetivo de maximizar su utilidad (medida de bienestar familiar) sujeto a restricciones presupuestarias. Desde la teoría del consumidor, las decisiones de los individuos implican la asignación de los recursos escasos entre consumo, ahorro, educación y trabajo. En este marco, las decisiones relacionadas con la inversión del capital humano como la educación y la capacitación representan elecciones racionales con el fin de incrementar sus ingresos en el futuro (Becker 1964). La literatura que analiza los determinantes del ingreso se divide en dos grandes grupos: el capital humano y la demanda.

Para el primero, el ingreso de los individuos está determinado por múltiples factores que interactúan entre sí en diferentes niveles. A partir de un enfoque microeconómico la teoría del capital humano sostiene que la educación y la experiencia constituyen variables fundamentales en la dinámica del ingreso de los hogares Becker (1964) y Mincer (1974). Esta teoría considera que existe una correlación positiva entre la escolaridad y la experiencia laboral con el ingreso. Por lo anterior, a mayor escolaridad, los individuos esperan adquirir habilidades más especializadas, generando mayores ingresos. De la misma manera, la experiencia laboral permite desarrollar competencias que se traducen en mayores remuneraciones. Asimismo, Mincer (1974) plantea un modelo de regresión lineal en el que vincula el logaritmo natural del ingreso con el nivel de escolaridad y la experiencia laboral. La teoría del capital humano propone una ecuación que relaciona el nivel de ingreso a partir de las inversiones que realizan los individuos en la formación de capacidades y habilidades (Willis, 1986).

Por otro lado, el enfoque del lado de la demanda enfatiza en los elementos asociados a la actividad económica, empleo, estructuras de mercado de trabajo y características de la empresa. Este enfoque critica el modelo de Mincer (1974), impulsando diferentes propuestas metodológicas que cuestionaron la validez del modelo. De acuerdo con Gaviria et al., (2007)

dicha ecuación no es aplicable a distintos contextos económicos, ya que no considera factores que determina el ingreso como el género, el nivel socioeconómico, escolaridad de los padres, el tamaño de la empresa donde labora el individuo y rasgos étnicos-culturales que pueden estar asociados a sesgos en el mercado laboral.

Por otro lado, Muñoz (2004) considera que existen otras características como los rasgos del jefe de familia, la composición generacional y de género influyen en el ingreso. Los factores como el acceso a la educación, la informalidad laboral y diferencias regionales desempeñan un rol fundamental en la determinación del ingreso. En este sentido, el contexto socioeconómico condiciona las oportunidades laborales y amplía la heterogeneidad de los hogares.

En México, diferentes autores han estudiado los determinantes del ingreso de los hogares. En Zamudio (1995), se emplea la información de la Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH) para estimar una ecuación de ingresos por medio de mínimos cuadrados ordinarios (MCO) y máxima verosimilitud (MV). Los resultados arrojaron que los individuos que continúan con los estudios superiores provienen de un hogar con menores integrantes y con un jefe de familia con mayores estudios en promedio.

Para Guzmán et al., (2005) las familias con una mejor calidad de capital humano están en condiciones de tener mayor capacidad en la asignación de sus miembros en el mercado laboral, potenciando sus posibilidades de mejorar su nivel de ingresos. El trabajo de Griliches (1977) señala que la escolaridad es un elemento importante en la determinación del ingreso, concluyendo que es necesario vincular las relaciones entre la academia y las empresas. En este sentido, Yúnez y Taylor (2004) presentan que la educación promueve la productividad y la competencia para erradicar la pobreza. No obstante, la desigualdad respecto al acceso a la educación exacerba la disparidad de ingresos entre regiones. La existencia de dos principales enfoques ha extendido las investigaciones en México a nivel regional, estatal o municipal a partir de microdatos relacionados a información sobre los hogares por medio de métodos econométricos tradicionales (Varela et al., 2010).

Segunda Sección: Marco Metodológico

En esta sección discutiremos lo referente a el modelo inferencial con el cual realizamos nuestro análisis.

2.1. Mínimos Cuadrados Ordinarios (MCO)

El modelo base para la estimación se formula mediante una regresión lineal múltiple bajo el supuesto clásico de exogeneidad de las variables explicativas. Sea y_i el ingreso corriente del hogar i , se especifica como:

$$y_i = x_i' \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

donde x_i representa un vector de características observables del hogar y del jefe del hogar, como el nivel educativo, el tamaño del hogar o el número de ocupados. Por otro lado, β es

el vector de parámetros a estimar y ε_i es el término de error. En un primer momento, el modelo se estima mediante Mínimos Cuadrados Ordinarios (OLS por sus siglas en inglés), técnica que proporciona estimadores insesgados bajo el supuesto de independencia entre x_i y ε_i .

No obstante, este supuesto se ve comprometido si alguna de las variables explicativas es endógena. En este caso, se argumenta que la participación del jefe del hogar en el sector formal podría no ser exógena, dado que decisiones como cotizar o no al sistema formal están correlacionadas con variables no observadas que también afectan el ingreso.

2.2 Modelo de dos etapas (Heckman)

Para corregir el sesgo derivado de la endogeneidad y posible selección muestral, se recurre al modelo de corrección propuesto por Heckman (1979). Esta metodología considera que la observación del ingreso está condicionada a una decisión previa: trabajar en el sector formal. Esta decisión puede estar influida por factores como la educación, la edad o la ocupación.

El modelo de Heckman se estima en dos etapas. En la primera, se estima una ecuación de selección mediante un modelo Probit:

$$s_i^* = z_i' \gamma + u_i, \quad s_i = 1 \quad \text{si} \quad s_i^* > 0$$

donde si s_i^* es una variable indicadora que toma el valor de 1 si el jefe del hogar trabaja en el sector formal. El vector z_i incluye variables que explican esta decisión, algunas de las cuales pueden no estar incluidas en x_i .

En la segunda etapa (ecuación de resultados), se estima el ingreso condicional a la selección con la inclusión de la *inversa de Mills*, calculada como $\lambda_i = \frac{\phi(z_i' \hat{\gamma})}{\Phi(z_i' \hat{\gamma})}$ donde $\phi(o)$ y $\Phi(o)$ son la función de densidad y distribución acumulada estándar normal, respectivamente.

El modelo corregido por sesgo de selección de participar en el mercado laboral formal es:

$$y_i = x_i' \beta + \rho \sigma_u \lambda_i + \eta_i$$

donde $\rho \sigma_u \lambda_i$ representa el término de corrección por sesgo de selección. Si el coeficiente de λ_i es significativo, se confirma que el sesgo de selección no puede ser ignorado (Vella, 1998).

2.3. Machine Learning LASSO

El objetivo principal de este estudio es identificar y analizar el papel que desempeñan las características individuales de los hogares, considerando su impacto diferenciado a través de los distintos deciles de ingreso y controlado el sesgo de autoselección de participar en el mercado laboral formal en México. Para llevar a cabo una selección robusta de variables explicativas, se implementó el algoritmo **Least Absolute Shrinkage and Selection Operator** (LASSO). Este método, ampliamente reconocido en la literatura econométrica y de machine learning, permite seleccionar las variables más relevantes al aplicar un factor de

penalización, conocido como λ , que reduce a cero los coeficientes asociados a los regresores con menor importancia relativa (Cameron & Trivedi, 2022). De este modo, LASSO no solo facilita la identificación de predictores significativos, sino que también minimiza el riesgo de sobreajuste en los modelos y mitiga problemas derivados de la multicolinealidad.

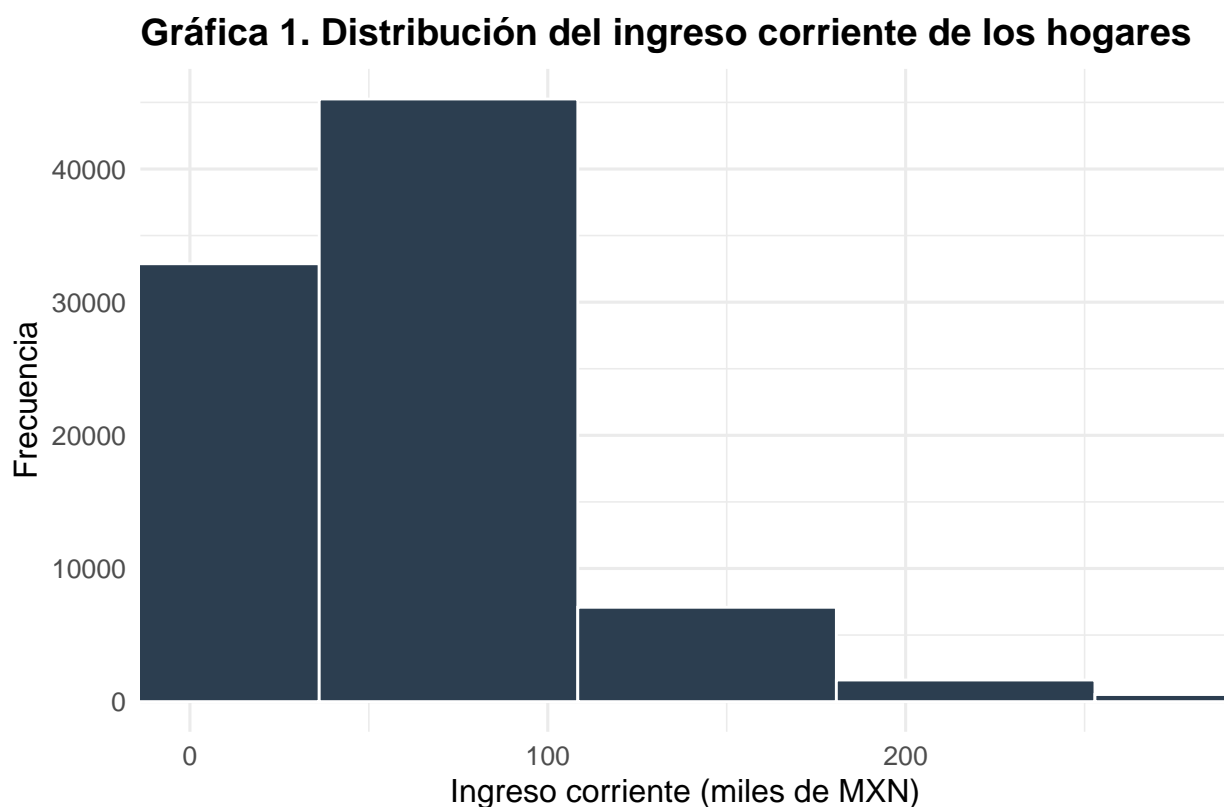
En el presente análisis, la variable dependiente de la ecuación de selección corresponde a si el hogar pertenece al sector formal de trabajo $formal_i$. Por otro lado, la variable dependiente de la ecuación de resultados corresponde al logaritmo natural del ingreso corriente por hogar ($\ln(ing_cor)$). Inicialmente, el modelo incorporó un conjunto amplio de 24 predictores, que representan tanto las características socioeconómicas de los hogares como aspectos específicos de su estructura y composición. Una vez aplicado el algoritmo, se descartaron aquellas variables cuyos coeficientes fueron llevados a cero, lo que indica una falta de impacto sustantivo en la explicación de las variaciones en la variable dependiente.

Tercera Sección: Hechos estilizados

Esta sección presenta un conjunto de hechos estilizados derivados del análisis exploratorio de los datos provenientes de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) del año 2022 elaborada por el Instituto Nacional de Estadística y Geografía (INEGI). A través de representaciones gráficas y estadísticas descriptivas, se exponen regularidades empíricas que anticipan los patrones estructurales en la distribución del ingreso. Aunque no se realiza inferencia causal en esta sección, estas regularidades permiten formular expectativas razonables sobre el comportamiento de las variables explicativas en el modelo econométrico posterior.

3.1 Distribución del ingreso corriente

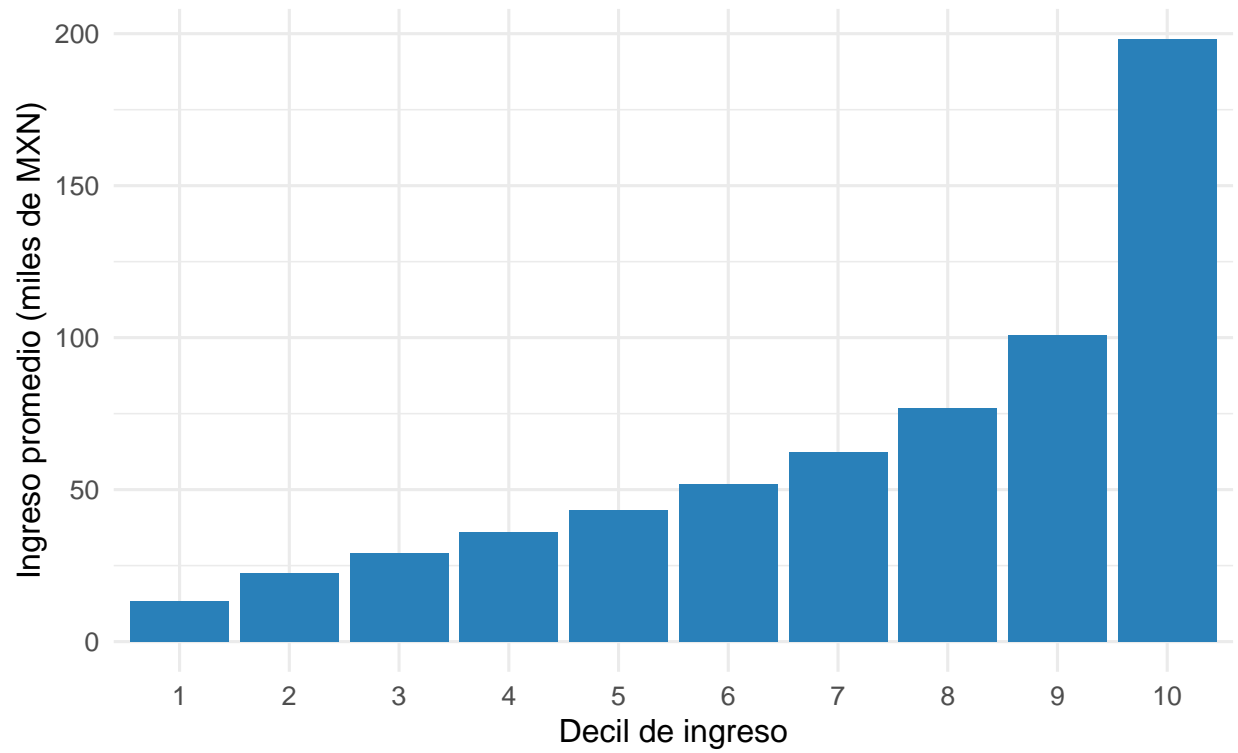
La Gráfica 1 revela una clara asimetría positiva en la distribución del ingreso corriente de los hogares. La fuerte concentración en los valores bajos y la cola alargada hacia ingresos altos son indicativos de un patrón regresivo, común en economías con niveles elevados de informalidad y baja redistribución fiscal (Deininger & Squire, 1996).



Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

Por su parte, la Gráfica 2 ilustra un crecimiento acelerado del ingreso conforme se asciende en los deciles de la distribución, sugiriendo una fuerte desigualdad en la base muestral. La disparidad entre el decil 10 y el resto de los grupos evidencia un fenómeno de concentración económica en la cúspide, alineado con hallazgos de López-Calva y Lustig (2010) sobre la persistencia de la desigualdad en América Latina.

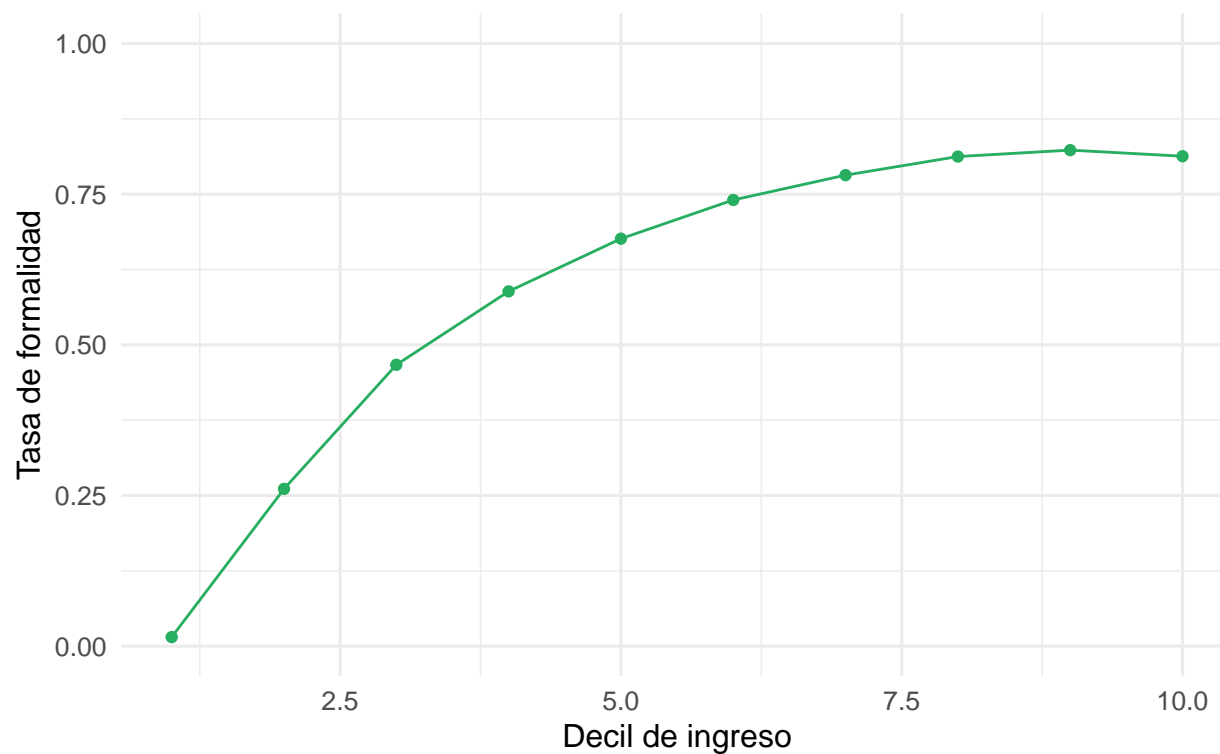
Gráfica 2. Ingreso corriente promedio por decil



Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

3.2 Tasa de formalidad por decil

Gráfica 3. Tasa de formalidad laboral por decil de ingreso

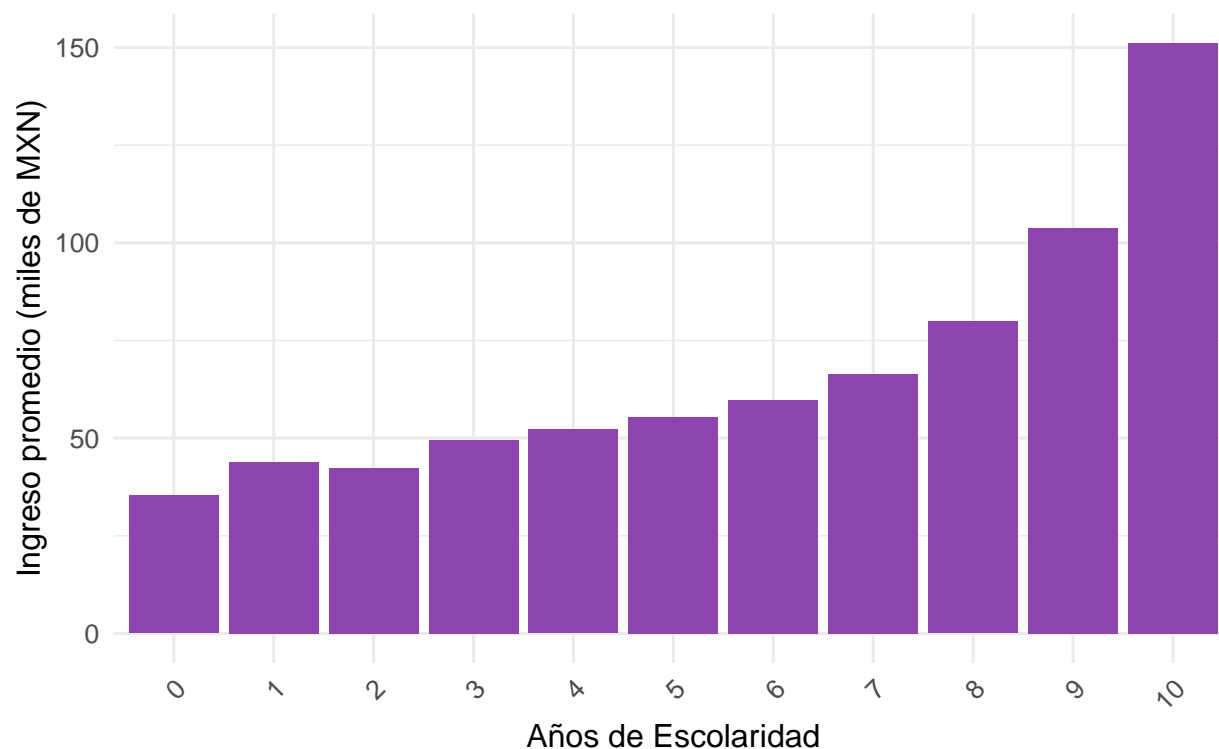


Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

Asimismo, en la Gráfica 3 se observa una relación directa entre ingreso y formalidad laboral. Este patrón sugiere que el acceso a empleos formales no solo es escaso en los deciles bajos, sino también dependiente de otras barreras estructurales. Esta observación empírica está en consonancia con el enfoque de exclusión y segmentación laboral discutido por Perry et al. (2007).

3.3 Ingreso promedio por nivel educativo del jefe del hogar

Gráfica 4. Ingreso promedio según escolaridad del jefe del hogar

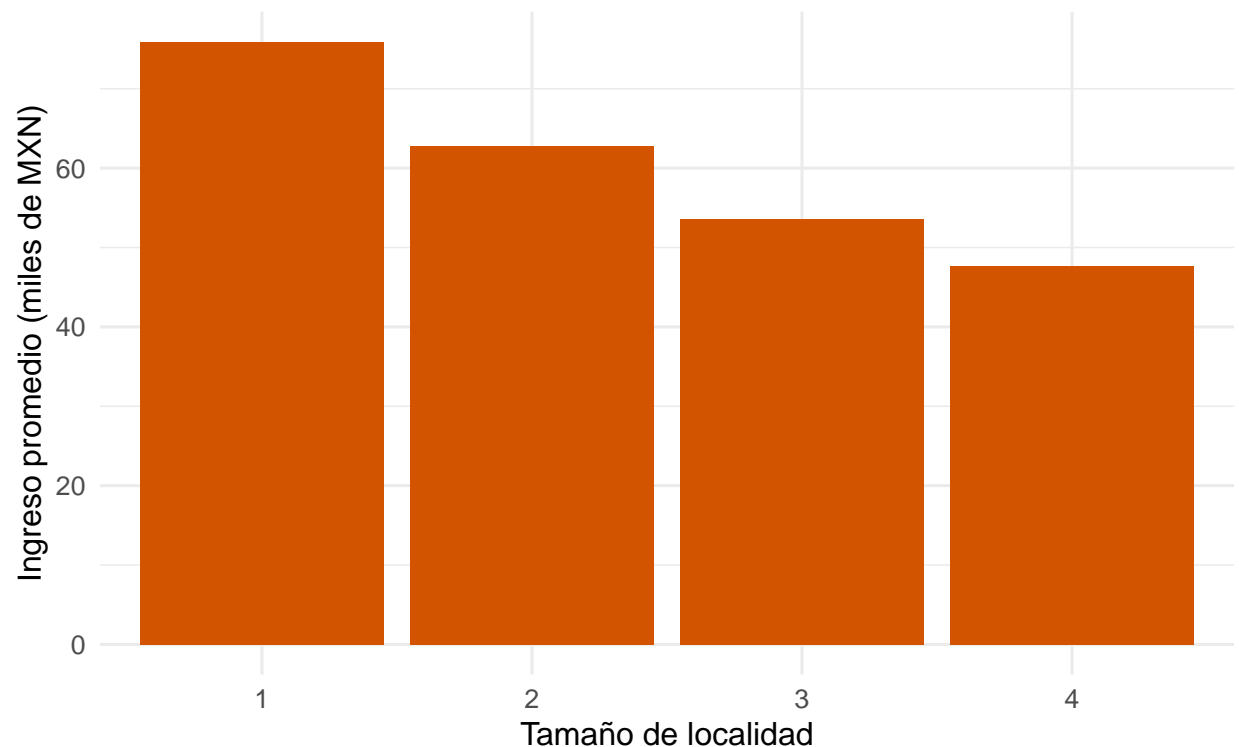


Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

Por su parte, la Gráfica 4 muestra un incremento consistente del ingreso conforme aumenta la escolaridad del jefe del hogar. Este resultado es coherente con la teoría del capital humano (Becker, 1964) y estudios posteriores como los de Psacharopoulos y Patrinos (2018), que cuantifican el rendimiento económico de la educación en contextos emergentes.

3.4 Ingreso y tamaño de localidad

Gráfica 5. Ingreso promedio por tamaño de localidad



Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

De la misma forma, la Gráfica 5 sugiere que los hogares ubicados en localidades urbanas mayores presentan mayores niveles de ingreso corriente, en comparación con aquellos asentados en localidades rurales. Esto puede explicarse por la concentración de empleos formales y mejor remunerados en contextos urbanos, fenómeno señalado por Glaeser y Gottlieb (2009).

3.5 Estadísticas descriptivas

Table 1: Estadísticas descriptivas de variables seleccionadas

	Variable	Media	Desv. Est.	Mín	Mediana	Máx
ing_cor	Ingreso corriente	60715.17	77430.61	405	45516.38	7153770
edad_jefe	Edad del jefe del hogar	51.22	15.92	13	50.00	109
tot_integ	Total de integrantes del hogar	3.35	1.64	1	3.00	10
ocupados	Número de ocupados del hogar	1.62	1.03	0	1.00	5
formal	El hogar participa en el sector formal	0.59	0.49	0	1.00	1

Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI.

La Tabla 1 resume las características principales de las variables de interés. Se aprecia una dispersión elevada en el ingreso corriente y una proporción significativa de ocupación informal. Estos indicadores sugieren la existencia de heterogeneidad estructural en el mercado laboral mexicano que será relevante para el análisis econométrico posterior. En otras palabras, los hogares se autoseleccionaron para participar en el mercado laboral formal, esta cuestión será discutida con detenimiento en la sección posterior.

Cuarta Sección: Resultados

4.1 Resultados de modelos OLS

La Tabla 2 presenta los resultados de tres modelos estimados mediante OLS, con el objetivo de identificar los principales determinantes del ingreso. El **Modelo 0**, que incluye únicamente el término constante, ofrece una línea base carente de poder explicativo (R^2 de 0.000), lo que evidencia la necesidad de introducir variables explicativas que capturen la heterogeneidad observada en los hogares.

Table 2: Modelos de regresión

Variable	Modelo 0	Modelo 1	Modelo 2
Constante	10.723*** (0.003)	8.883*** (0.02)	11.999*** (0.007)
Sexo jefe hogar		0.042*** (0.005)	0.006*** (0.001)
Edad jefe		0.009*** (0.001)	0 (0)
Edad jefe ²		0 (0)	0 (0)
Educacion jefe		0.131*** (0.001)	0.006*** (0)
Clase hogar		0.133*** (0.004)	0.007*** (0.001)
Total integrantes		0.028*** (0.002)	0.001 (0.001)
Ocupados		0.233*** (0.003)	0.004*** (0.001)
Hombres		0.021*** (0.003)	-0.001 (0.001)
Trab agricolas		-0.452*** (0.007)	-0.021*** (0.002)
Trab industriales		-0.163*** (0.009)	-0.005* (0.003)
Trab comercio		-0.131*** (0.007)	0.002 (0.002)
Trab servicios		-0.11*** (0.007)	-0.004* (0.002)
Decil 1			-2.608*** (0.003)
Decil 2			-2.043*** (0.003)
Decil 3			-1.78*** (0.003)
Decil 4			-1.575*** (0.003)
Decil 5			-1.39*** (0.003)
Decil 6			-1.213*** (0.003)
Decil 7			-1.031*** (0.003)
Decil 8			-0.827*** (0.003)

Decil 9			-0.562*** (0.003)
R-cuadrado	0	0.37	0.942
AIC	197312.723	156668.976	-53779.033
Observaciones	87951	87951	87951

*Nota: Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$* Fuente: Elaboración propia con datos

En el **Modelo 1**, se incorporan variables sociodemográficas y laborales del jefe del hogar y de los miembros del hogar. Este modelo explica aproximadamente el 37% de la variabilidad en el ingreso, lo cual representa una mejora sustancial respecto al modelo base. Las estimaciones muestran que pertenecer al sector formal del mercado laboral se asocia con un ingreso significativamente mayor, con un coeficiente positivo y altamente significativo. Asimismo, el sexo del jefe del hogar, su edad, la estructura del hogar, y especialmente su nivel educativo, tienen efectos estadísticamente significativos sobre el ingreso. Estas relaciones se alinean con la literatura económica, que documenta consistentemente una prima salarial asociada a la formalidad y a mayores niveles educativos.

El **Modelo 2** profundiza en este análisis al incorporar los deciles de ingreso como variables explicativas, lo cual permite capturar no solo las características individuales y del hogar, sino también la posición relativa del hogar dentro de la distribución del ingreso. La inclusión de estas variables mejora dramáticamente el ajuste del modelo (R^2 de 0.943), indicando que la posición en la estructura distributiva es altamente informativa para explicar las diferencias de ingreso. No obstante, algunos coeficientes previamente significativos en el Modelo 1 pierden magnitud o significancia estadística, lo cual sugiere una potencial colinealidad o mediación a través de la posición en el ingreso.

No obstante, es importante señalar que la estimación mediante MCO podría estar sesgada debido a problemas de endogeneidad, particularmente en relación con la variable de participación en el sector formal. Esta condición no es asignada aleatoriamente, sino que puede estar influida por características no observadas como la habilidad, la motivación o las redes sociales del jefe del hogar, que también inciden directamente en el ingreso. En este sentido, existe un riesgo claro de **autoselección**, ya que los hogares no participan en el sector formal de manera exógena, sino que podrían seleccionar su estatus laboral en función de atributos no observables correlacionados con el ingreso.

Este posible sesgo de selección constituye una limitación metodológica crítica para la interpretación causal de los coeficientes estimados. Por tanto, se justifica la estimación de modelos que corrijan explícitamente por esta fuente de endogeneidad. En particular, los modelos presentados más adelante (Modelos 3 y 4) utilizan una estrategia de corrección por selección muestral al estilo de Heckman, lo cual permite abordar de manera explícita esta preocupación empírica, incorporando una ecuación de selección que modela la probabilidad de participar en el sector formal. Con ello, se busca obtener estimaciones insesgadas del efecto de la formalidad y de las demás covariables sobre el ingreso, corrigiendo el sesgo que pudiera derivarse de una muestra no aleatoria de observaciones.

4.2 Resultados de modelo Heckman

Por lo anterior, la Tabla 3 presenta los resultados del Modelo 3, estimado mediante el método de Heckman de dos etapas. Esta especificación permite corregir el sesgo de selección muestral derivado de la decisión no aleatoria de los hogares de participar en el sector formal, un problema que puede generar endogeneidad en las estimaciones previas, al estar restringidas a la subpoblación observada.

```
## [1] "Tabla 3. Modelo Heckman"
```

```
## -----
```

```
## Tobit 2 model (sample selection model)
```

```
## Maximum Likelihood estimation
```

```
## Newton-Raphson maximisation, 4 iterations
```

```
## Return code 2: successive function values within tolerance limit (tol)
```

```
## Log-Likelihood: -10259.68
```

```
## 87951 observations (36466 censored and 51485 observed)
```

```
## 38 free parameters (df = 87913)
```

```
## Probit selection equation:
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-1.456e+00	5.599e-02	-26.001	< 2e-16 ***
## sexo_jh	1.141e-01	1.233e-02	9.250	< 2e-16 ***
## edad_jefe	7.686e-03	2.027e-03	3.792	0.000149 ***
## edad_jh2	-2.077e-04	1.902e-05	-10.919	< 2e-16 ***
## educa_jefe	9.381e-02	2.297e-03	40.845	< 2e-16 ***
## clase_hog	1.551e-01	1.216e-02	12.761	< 2e-16 ***
## tot_integ	3.627e-02	7.506e-03	4.832	1.35e-06 ***
## ocupados	8.739e-01	8.247e-03	105.959	< 2e-16 ***
## hombres	6.046e-02	8.201e-03	7.373	1.69e-13 ***
## trab_agro	-1.659e+00	1.925e-02	-86.155	< 2e-16 ***
## trab_ind	-1.263e+00	2.257e-02	-55.953	< 2e-16 ***
## trab_com	-1.205e+00	1.898e-02	-63.506	< 2e-16 ***
## trab_serv	-1.264e+00	1.775e-02	-71.223	< 2e-16 ***
## menores	-7.498e-02	8.914e-03	-8.411	< 2e-16 ***

```
## Outcome equation:
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	1.200e+01	8.552e-03	1403.364	< 2e-16 ***
## sexo_jh	6.771e-03	1.508e-03	4.491	7.09e-06 ***
## edad_jefe	2.318e-04	2.615e-04	0.886	0.3754
## edad_jh2	5.413e-06	2.603e-06	2.079	0.0376 *
## educa_jefe	5.922e-03	3.238e-04	18.292	< 2e-16 ***
## clase_hog	-1.070e-03	1.552e-03	-0.689	0.4907
## tot_integ	-5.998e-04	6.976e-04	-0.860	0.3899
## ocupados	2.572e-05	1.110e-03	0.023	0.9815

```

## hombres      6.798e-06  9.081e-04   0.007   0.9940
## trab_agro    1.773e-02  3.576e-03   4.959  7.10e-07 ***
## trab_ind     1.814e-03  3.541e-03   0.512   0.6085
## trab_com     6.898e-03  2.892e-03   2.385   0.0171 *
## trab_serv    4.678e-03  2.710e-03   1.726   0.0843 .
## decil1      -2.255e+00  1.198e-02 -188.192 < 2e-16 ***
## decil2      -1.996e+00  3.690e-03 -541.023 < 2e-16 ***
## decil3      -1.756e+00  3.085e-03 -569.231 < 2e-16 ***
## decil4      -1.555e+00  2.881e-03 -539.570 < 2e-16 ***
## decil5      -1.370e+00  2.743e-03 -499.269 < 2e-16 ***
## decil6      -1.194e+00  2.651e-03 -450.172 < 2e-16 ***
## decil7      -1.012e+00  2.580e-03 -392.136 < 2e-16 ***
## decil8      -8.087e-01  2.521e-03 -320.791 < 2e-16 ***
## decil9      -5.433e-01  2.492e-03 -218.033 < 2e-16 ***
##      Error terms:
##              Estimate Std. Error t value Pr(>|t|)
## sigma  0.1415899   0.0004588 308.628 < 2e-16 ***
## rho    -0.1305619   0.0164978  -7.914 2.52e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## -----

## [1] "Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI."

```

En la ecuación de selección, los resultados muestran que la probabilidad de participar en el sector formal se ve incrementada por características individuales y del hogar. En particular, el coeficiente estimado para sexo_jh (0.114) indica que los hogares encabezados por hombres tienen una mayor probabilidad de inserción formal respecto a sus contrapartes lideradas por mujeres. La edad del jefe del hogar presenta una relación no lineal, con un efecto positivo en su forma lineal (0.0077) y negativo en el término cuadrático (-0.00021), lo que sugiere que la probabilidad de formalidad aumenta con la edad pero a un ritmo decreciente. La educación también incide positivamente sobre dicha probabilidad: cada año adicional de escolaridad del jefe del hogar incrementa en 0.094 puntos la probabilidad de participar en el sector formal, lo cual es consistente con teorías de segmentación laboral y acumulación de capital humano.

Asimismo, el número de ocupados en el hogar (0.874) tiene un efecto considerable, lo que podría reflejar tanto redes familiares de inserción laboral como economías de escala asociadas a hogares con mayor participación en el mercado. Por otro lado, las variables que capturan el sector de actividad económica (trab_agro, trab_ind, trab_com, trab_serv) muestran coeficientes negativos de alta magnitud y significancia estadística, lo que sugiere que los hogares cuya ocupación principal se ubica en sectores tradicionalmente informales enfrentan barreras estructurales para la incorporación al sector formal. Por ejemplo, trab_agro exhibe un coeficiente de -1.659, lo que implica una probabilidad sustancialmente menor de formalidad respecto al sector de referencia.

En la ecuación de resultado, que modela el ingreso corriente condicionado a la participación formal y corrige el sesgo de selección, se observa que el efecto de muchas variables se modifica notablemente respecto a las estimaciones anteriores. La variable `sexo_jh` conserva un efecto positivo y estadísticamente significativo (0.0068), lo que sugiere que, entre los hogares que participan formalmente, los encabezados por hombres tienden a obtener ingresos ligeramente mayores. El nivel educativo del jefe del hogar continúa siendo una de las variables más relevantes, con un coeficiente de 0.0059, lo que confirma la existencia de retornos significativos a la educación incluso después de corregir por la selección.

Las variables que capturan la actividad económica también muestran efectos diferenciados. Por ejemplo, `trab_agro` presenta un coeficiente positivo y significativo (0.0177), lo cual contrasta con su efecto en la ecuación de selección y sugiere que, dentro del sector formal, los hogares vinculados al agropecuario perciben ingresos relativamente altos. De manera similar, `trab_com` y `trab_serv` muestran coeficientes positivos (0.0069 y 0.0047, respectivamente), aunque de menor magnitud, indicando que la inserción formal en estos sectores está asociada con ingresos ligeramente superiores al sector base.

Por su parte, las variables de nivel de ingreso (`decil1` a `decil9`) mantienen un gradiente creciente y negativo, con coeficientes de gran magnitud y significancia, lo cual es coherente con la transformación logarítmica de la variable dependiente. Por ejemplo, el coeficiente de `decil1` es -2.255, lo que implica que, incluso entre los hogares formales, los ubicados en los deciles más bajos presentan ingresos sustancialmente menores que aquellos del decil de referencia.

Por otro lado, el parámetro de correlación entre los términos de error de ambas ecuaciones (ρ de -0.131) resulta estadísticamente distinto de cero, lo cual confirma la existencia de sesgo de selección. La correlación negativa implica que los hogares con características no observables que incrementan su probabilidad de participar en el sector formal tienden, en promedio, a tener menores ingresos que aquellos que no participan, lo que valida empíricamente la necesidad de aplicar una estrategia de corrección por selección muestral.

4.3 Resultado de modelo Heckman con LASSO

La Tabla 4 muestra los resultados del ajustado mediante regularización LASSO.¹ Esta aproximación mejora la parsimonia del modelo y ayuda a mitigar problemas de sobreajuste, especialmente en contextos con alta dimensionalidad o multicolinealidad entre predictores.

```
## [1] "Tabla 4. Modelo Heckman con LASSO"
```

```
## -----
```

```
## Tobit 2 model (sample selection model)
```

```
## Maximum Likelihood estimation
```

```
## Newton-Raphson maximisation, 4 iterations
```

```
## Return code 8: successive function values within relative tolerance limit (reltol)
```

¹Véase Tabla A.1 Resultados del LASSO.


```

## Log-Likelihood: -10638.02
## 87951 observations (36466 censored and 51485 observed)
## 32 free parameters (df = 87919)
## Probit selection equation:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.7077115  0.0312952  -22.61  < 2e-16 ***
## edad_jefe   -0.0116454  0.0003535  -32.95  < 2e-16 ***
## hombres     0.0659885  0.0062771   10.51  < 2e-16 ***
## ocupados    0.9174537  0.0075648  121.28  < 2e-16 ***
## trab_agro   -1.6716439  0.0191459  -87.31  < 2e-16 ***
## trab_ind    -1.2632265  0.0225187  -56.10  < 2e-16 ***
## trab_com    -1.2142230  0.0189104  -64.21  < 2e-16 ***
## trab_serv   -1.2650895  0.0177048  -71.45  < 2e-16 ***
## clase_hog1  -0.2607070  0.0168797  -15.45  < 2e-16 ***
## jefe_mujer  -0.0923949  0.0119843   -7.71  1.27e-14 ***
## edu_jefe     0.1866040  0.0044108   42.31  < 2e-16 ***
## Outcome equation:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.085e+01  3.930e-03 2759.674  < 2e-16 ***
## sexo_jh      5.529e-03  1.442e-03   3.833 0.000127 ***
## edad_jefe    5.287e-04  5.042e-05  10.486  < 2e-16 ***
## trab_agro    2.313e-02  3.407e-03   6.791 1.12e-11 ***
## trab_ind     5.836e-03  3.478e-03   1.678 0.093324 .
## trab_com     1.240e-02  2.813e-03   4.406 1.05e-05 ***
## trab_serv    1.137e-02  2.607e-03   4.360 1.30e-05 ***
## ocupados    -7.131e-03  9.502e-04  -7.505 6.21e-14 ***
## clase_hog1   8.048e-03  2.947e-03   2.731 0.006309 **
## clase_hog2   5.995e-03  1.561e-03   3.841 0.000123 ***
## decil1      -1.068e+00  1.193e-02 -89.530  < 2e-16 ***
## decil2      -8.101e-01  3.421e-03 -236.769  < 2e-16 ***
## decil3      -5.680e-01  2.809e-03 -202.231  < 2e-16 ***
## decil4      -3.648e-01  2.638e-03 -138.269  < 2e-16 ***
## decil5      -1.778e-01  2.541e-03  -69.961  < 2e-16 ***
## decil7       1.837e-01  2.470e-03  74.376  < 2e-16 ***
## decil8       3.895e-01  2.461e-03 158.252  < 2e-16 ***
## decil9       6.586e-01  2.487e-03 264.789  < 2e-16 ***
## decil10      1.208e+00  2.541e-03 475.540  < 2e-16 ***
## Error terms:
##           Estimate Std. Error t value Pr(>|t|)
## sigma    0.1428591  0.0004702 303.80  <2e-16 ***
## rho     -0.2017428  0.0120299  -16.77  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## -----

```

[1] "Fuente: Elaboración propia con datos de la ENIGH 2022, INEGI."

En la ecuación de selección, los coeficientes estimados revelan que ciertas variables mantienen una influencia significativa en la probabilidad de participar en el sector formal. La edad del jefe del hogar muestra un efecto negativo (-0.0116), lo que sugiere que, bajo este ajuste, el avance etario se asocia con una menor probabilidad de formalidad, en contraste con la especificación previa. Este cambio puede atribuirse a la penalización introducida por el LASSO, que ajusta la magnitud de los coeficientes para minimizar la función objetivo.

La variable hombres conserva un efecto positivo, reafirmando que los hogares con mayor presencia masculina enfrentan menores barreras de entrada al empleo formal. El número de ocupados en el hogar continúa siendo altamente relevante, lo que enfatiza la importancia de las dinámicas familiares en la inserción laboral. Los coeficientes de los sectores económicos permanecen negativos y significativos, con valores similares a los del modelo sin regularización, lo cual refuerza la evidencia de segmentación estructural del mercado laboral mexicano.

Se destaca también la inclusión de variables como `clase_hog1` y `jefe_mujer`, que reflejan obstáculos adicionales en la selección hacia la formalidad para hogares monoparentales y aquellos encabezados por mujeres, respectivamente. La educación del jefe del hogar mantiene un efecto robusto, lo que subraya nuevamente el papel del capital humano en la transición hacia el sector formal.

En la ecuación de resultado, que estima el ingreso corriente condicionado a la participación formal, los efectos seleccionados por LASSO sugieren patrones similares a los modelos anteriores, aunque con algunas diferencias en magnitudes y significancia. El coeficiente de `sexo_jh` confirma una ligera ventaja de ingresos para los hogares encabezados por hombres. La edad tiene ahora un efecto positivo, indicando que, entre los formalmente insertados, el envejecimiento se asocia con ingresos mayores, consistente con trayectorias laborales más consolidadas.

Las variables de sector económico muestran efectos positivos, especialmente en `trab_agro` y `trab_com`, aunque `trab_ind` es marginalmente significativo. En contraste con su efecto negativo en la ecuación de selección, estas estimaciones sugieren que, una vez dentro del sector formal, estas actividades pueden ofrecer ingresos comparables o incluso superiores al sector base.

En cuanto a los deciles de ingreso, la pendiente de los coeficientes es coherente con una transformación logarítmica: los hogares en los deciles más bajos (`decil1` a `decil5`) presentan coeficientes negativos significativos, mientras que a partir del `decil7` se observan valores positivos crecientes, alcanzando 1.208 para `decil10`. Esto sugiere que el modelo logra capturar adecuadamente la dispersión del ingreso entre los distintos estratos socioeconómicos.

Finalmente, el parámetro de correlación de los errores (ρ de -0.202) mantiene un valor negativo y significativo, reafirmando la existencia de sesgo de selección no aleatoria. La magnitud superior respecto al modelo sin regularización podría indicar que el LASSO ha eliminado variables que actuaban como atenuantes de esta correlación, dejando en evidencia un vínculo más claro entre la decisión de participar en el sector formal y el ingreso observado.

Quinta Sección: Discusión

Los resultados obtenidos en las Tablas 3 y 4, a partir de la estimación de modelos Heckman con y sin regularización LASSO, ofrecen evidencia robusta sobre los factores que condicionan la inserción laboral formal y el nivel de ingreso en los hogares mexicanos, utilizando microdatos de la ENIGH 2022.

Primero, la presencia de un sesgo de selección significativo (evidenciado por el coeficiente negativo y significativo de correlación entre los errores de las ecuaciones de selección y resultado ρ) confirma que la participación en el sector formal no es aleatoria y está condicionada por características observables y no observables del hogar. Este hallazgo es consistente con la literatura que advierte sobre la necesidad de corregir sesgos en la estimación de ingresos laborales para evitar inferencias erróneas (Heckman, 1979; Wooldridge, 2010).

En cuanto a la selección, variables como la educación del jefe del hogar, el género y la composición laboral del hogar emergen como factores determinantes. La educación, cuyo efecto positivo es robusto en ambos modelos, refuerza la tesis de que el capital humano es un determinante clave en el acceso a empleos formales con mejores condiciones (Mincer, 1974; Psacharopoulos, 1994). La evidencia de que hogares encabezados por mujeres o con mayor proporción de trabajadores en sectores informales presentan menor probabilidad de formalidad es congruente con estudios previos que documentan desigualdades estructurales y discriminación en el mercado laboral mexicano (Oaxaca, 1973; Mora, 2011).

La implementación del LASSO para la selección de variables en el Modelo 4 mejora la parsimonia sin sacrificar la significancia de los predictores principales, lo cual confirma la utilidad de esta técnica para manejar la alta dimensionalidad y reducir la multicolinealidad, alineándose con estudios metodológicos recientes (Tibshirani, 1996; Belloni et al., 2014).

En la ecuación de resultados, la distribución del ingreso muestra un gradiente claro entre deciles, lo que refleja la persistencia de desigualdades económicas profundas en México, conforme a lo señalado por López y Rodríguez (2019). Asimismo, los efectos sectoriales en el ingreso formal sugieren que, aunque ciertos sectores tienen menores probabilidades de formalidad, una vez formalizados, los ingresos en actividades agropecuarias y comerciales pueden ser competitivos, lo que evidencia heterogeneidad interna en los sectores económicos (González y Mendoza, 2017).

Finalmente, el signo negativo del coeficiente ρ indica que los hogares con mayor propensión a no estar formalizados tienden a reportar ingresos más bajos, validando el enfoque de corrección de selección y reafirmando la importancia de políticas públicas que promuevan la formalización para mejorar las condiciones económicas de los hogares vulnerables (Heckman y Navarro-Lozano, 2004).

Conclusión

Los resultados empíricos derivados de la estimación de modelos de selección tipo Heckman, tanto con especificación tradicional como con regularización mediante LASSO, ofrecen ev-

idencia robusta sobre la existencia de mecanismos de autoselección en la inserción laboral formal de los hogares mexicanos, así como sobre la estructura determinante de sus niveles de ingreso. La presencia de un coeficiente de correlación significativo y negativo entre los términos de error de las ecuaciones de selección y de resultado (ρ) confirma la hipótesis de selección no aleatoria, lo cual valida la pertinencia metodológica del enfoque utilizado y subraya la necesidad de corregir dicho sesgo para obtener estimaciones no sesgadas de los determinantes del ingreso.

En la dimensión de selección al empleo formal, la educación del jefe del hogar emerge como un predictor consistente y significativo, reforzando la relevancia del capital humano en la estructuración de oportunidades laborales. Asimismo, características sociodemográficas como el género del jefe del hogar y la composición ocupacional del mismo evidencian patrones estructurales de exclusión y segmentación en el mercado de trabajo, particularmente en perjuicio de mujeres y trabajadores vinculados al sector informal.

El uso de regularización mediante LASSO en la especificación ampliada mejora la parsimonia del modelo sin pérdida de poder explicativo, lo que demuestra su utilidad como herramienta metodológica en contextos de alta dimensionalidad y posible multicolinealidad entre predictores. Este enfoque permite identificar un subconjunto más reducido y relevante de variables explicativas, optimizando la precisión de las inferencias.

En cuanto a los determinantes del ingreso laboral formal, los resultados reflejan un patrón persistente de desigualdad, expresado en la clara estratificación por deciles de ingreso, y revelan una heterogeneidad sectorial que, en algunos casos, compensa los menores niveles de formalidad con ingresos comparables o incluso competitivos, una vez que se logra la inserción formal.

En conjunto, estos hallazgos no solo confirman la importancia de abordar la informalidad laboral como un fenómeno estructural, sino que también respaldan la necesidad de políticas públicas orientadas a fortalecer la educación, reducir las barreras de acceso al empleo formal y promover la equidad de oportunidades en el mercado laboral. La correcta identificación de los factores que inciden en la formalización y en el nivel de ingresos permite avanzar hacia un diagnóstico más preciso y hacia intervenciones más eficaces en la reducción de la desigualdad económica en México.

Bibliografía

- Albert, C., & Ángeles, M. (2005). *Education, wages and job satisfaction*. Universidad de Castilla La Mancha (UCLM).
- Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.
- Bosch, M., & Campos-Vázquez, R. M. (2014). The trade-offs of social assistance programs in the labor market: The case of the ‘Seguro Popular’ program in Mexico. *American Economic Journal: Economic Policy*, 6(1), 71–99. <https://doi.org/10.1257/pol.6.1.71>

- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29–50. <https://doi.org/10.1257/jep.28.2.29>
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650. <https://doi.org/10.1093/restud/rdt044>
- Cabrera, C., Mungaray, A., Varela, R., & Hernández, E. (2008). Capital humano e ingresos en la manufactura de Tijuana y Mexicali: 1994–2001. *Estudios Fronterizos*, 9(18), 95–114.
- Cameron, A. C., & Trivedi, P. K. (2022). *Microeconometrics using Stata. Volume II: Nonlinear models and causal inference methods* (2nd ed.). Stata Press.
- Deininger, K., & Squire, L. (1996). A new data set measuring income inequality. *The World Bank Economic Review*, 10(3), 565–591. <https://doi.org/10.1093/wber/10.3.565>
- Fusco, A., & Islam, N. (2017). *Household size and poverty*. LISER.
- Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4), 983–1028. <https://doi.org/10.1257/jel.47.4.983>
- González, A., & Mendoza, R. (2017). Heterogeneidad sectorial y mercados laborales en México: una perspectiva microeconómica. *Revista Mexicana de Economía y Finanzas*, 12(1), 45–66.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, 45(1), 1–22. <https://doi.org/10.2307/1913285>
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161. <https://doi.org/10.2307/1912352>
- Heckman, J. J., Humphries, J. E., Veramendi, G., & Urzua, S. S. (2014). *Education, health and wages*. National Bureau of Economic Research.
- Heckman, J. J., & Landerso, R. (2022). Lessons for Americans from Denmark about inequality and social mobility. *Labour Economics*.
- López, M., & Rodríguez, C. (2019). Desigualdad y pobreza en México: análisis a partir de datos de ingreso y gasto. *Estudios Económicos*, 34(1), 75–102.
- López-Calva, L. F., & Lustig, N. (2010). *Declining inequality in Latin America: A decade of progress?* Brookings Institution Press.
- Mincer, J. (1974). *Schooling, experience, and earnings*. National Bureau of Economic Research.

- Mora, M. (2011). Mujeres y mercado laboral en México: análisis de brechas salariales y participación. *Revista Latinoamericana de Estudios Sociales*, 8(2), 105-129.
- Muluneh, D., & Addisu, Y. (2024). Determinants of urban household income diversification and its relation to food security: lessons from Yejube town in Ethiopia. *Cogent Social Sciences*.
- Muñoz, M. (2004). Determinantes del ingreso y del gasto corriente de los hogares. *Revista de Economía Institucional*, 6(10), 183-199.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693-709. <https://doi.org/10.2307/2525981>
- Perry, G., Arias, O., López, H., Maloney, W., & Servén, L. (2007). *Informality: Exit and exclusion*. World Bank Latin American and Caribbean Studies.
- Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, 22(9), 1325-1343. [https://doi.org/10.1016/0305-750X\(94\)90007-8](https://doi.org/10.1016/0305-750X(94)90007-8)
- Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, 26(5), 445-458. <https://doi.org/10.1080/09645292.2018.1484426>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Varela Llamas, R., Ocegueda Hernández, J. M., Castillo Ponce, R. A., & Huber Bernal, G. (2010). Determinantes de los ingresos salariales en México: Una perspectiva de capital humano. *Región y Sociedad*, 22(49), 117-142.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *The Journal of Human Resources*, 33(1), 127-169. <https://doi.org/10.2307/146317>
- Vu, V. H. (2020). The impact of education on household income in rural Vietnam. *International Journal of Financial Studies*.
- Wang, J., Xie, H., & Fisher, J. H. (2012). *Multilevel models: Applications using SAS*. De Gruyter.
- Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. En O. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics* (Vol. 1, pp. 525-602). Elsevier.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

- Yúnez Naude, A., & Taylor, J. E. (2004). Los determinantes de las actividades y el ingreso no agrícola de los hogares rurales de México, con énfasis en la educación. En *Empleo e ingresos rurales no agrícolas en América Latina* (pp. 231–245). Comisión Económica para América Latina y el Caribe (CEPAL). <https://hdl.handle.net/11362/6202>
- Zamudio, A. (1995). Rendimientos a la educación superior en México: Ajuste por sesgo utilizando máxima verosimilitud. *Economía Mexicana*, 4(1), 69–91.

Anexo

A.1 Resultados del LASSO

Table 3: Tabla A1. Ecuación de selección (Variable dependiente: formal)

Variable	Coefficiente
decil	0.020
sexo_jh	0.024
clase_hog	0.005
edad_jefe	-0.008
hombres	0.001
mujeres	-0.004
ocupados	0.122
trab_agro	-0.222
trab_ind	-0.217
trab_com	-0.205
trab_serv	-0.218
menores	0.003
sueldos	0.000
clase_hog1	-0.012
educa_jefe	0.006
edu_jefe	-0.018
edad_jh	0.199
edad_jh2	0.000
decil1	-0.291
decil2	-0.136
decil3	-0.007
decil4	0.046
decil5	0.072
decil6	0.070

decil7	0.041
decil9	-0.074
decil10	-0.259
tam_loc_1	0.014
tam_loc_2	0.011
tam_loc_4	-0.013

Fuente: Elaboración propia

Table 4: Tabla A2. Ecuación de resultados (Variable dependiente: (Log) ingreso corriente)

Variable	Coeficiente
formal	0.002
sexo_jh	0.000
edad_jefe	0.000
ocupados	0.002
trab_agro	-0.017
clase_hog1	-0.020
jefe_mujer	-0.001
edu_jefe	0.010
edad_jh2	0.000
decil1	-1.385
decil2	-0.822
decil3	-0.559
decil4	-0.354
decil5	-0.169
decil7	0.182
decil8	0.386
decil9	0.652
decil10	1.214
tam_loc_4	-0.003

Fuente: Elaboración propia