



Benemérita Universidad Autónoma de Puebla

Facultad de Economía

"Brecha Salarial de Género en México: Un Análisis Cuantitativo de la Influencia Socioeconómica (2019-2023)".

Tesis presentada para obtener el grado en Licenciado en Economía

Presenta: Steve Aguilar Hernández

Director de tesis: Dr. Juan Alberto Vázquez Muñoz

Diciembre, 2025

# "Brecha Salarial de Género en México: Un Análisis Cuantitativo de la Influencia Socioeconómica (2019-2023)"

Steve Aguilar Hernández

## Resumen

El presente trabajo examina la brecha salarial de género en México durante 2019-2023, período que abarca la irrupción y efectos diferenciados de la pandemia de COVID-19. Con microdatos de la ENOE, se caracteriza la participación laboral de mujeres y hombres y se documentan desigualdades persistentes incluso entre individuos con rasgos observables similares (escolaridad, edad, experiencia, ocupación y región). La tesis se inscribe en un enfoque cuantitativo y explicativo: se estiman ecuaciones salariales tipo Mincer, correcciones de tipo Heckman y descomposiciones Oaxaca-Blinder (doble y triple), complementadas con un modelo espacial SLX a nivel municipal. Este andamiaje permite separar la parte de la brecha asociada a diferencias en dotaciones de capital humano y estructura ocupacional, de aquella compatible con mecanismos de discriminación y segmentación del mercado de trabajo. Los resultados buscan aportar evidencia útil para el diseño de políticas orientadas a la equidad salarial y la reducción de desigualdades de género.

## Índice

<b>CAPÍTULO I. Introducción y marco teórico.....</b>	<b>5</b>
1.1 Introducción al problema .....	5
1.1.1 Justificación .....	7
1.1.2 Preguntas de investigación .....	9
1.1.3 Objetivo .....	10
1.1.4 Hipótesis .....	12
1.2 Revisión de literatura.....	13
1.2.1 Marco teórico general.....	13
1.2.2 Segregación ocupacional y brecha de género.....	18
1.2.3 Brecha salarial y estructura del mercado laboral mexicano .....	20
1.3 Marco metodológico general.....	23
<b>CAPÍTULO II. Estadística descriptiva y caracterización de la muestra .....</b>	<b>29</b>
2.1 Panorama estadístico .....	29
2.2 Comparación de la PEA.....	31
2.3 Primeras evidencias de brecha .....	36
<b>CAPÍTULO III. Metodología .....</b>	<b>43</b>
3.1 Observaciones técnicas .....	43
3.1.1 Construcción previa a los modelos, exclusión y conformación de las variables .....	43
3.1.2 Ponderador.....	46
3.2 Ecuación salarial (Mincer) por sexo .....	47
3.2 Mincer Base (por sexo) y Oaxaca Blinder.....	49
3.3 Variante con corrección de selección (Heckman) .....	52
3.4 Econometría espacial: modelo SLX .....	55
3.4.1 Estimaciones previas y generación de variables .....	55

3.4.2 Modelos SLX (spatial Lag o X).....	60
3.4.3 Implementación operativa .....	64
3.5 Consideraciones de identificación y limitación .....	66
<b>CAPÍTULO IV. Resultados .....</b>	<b>68</b>
4.1 Estimación Oaxaca Blinder.....	68
4.2 Oaxaca Blinder descomposición triple .....	77
4.3 Oaxaca Blinder con corrección por selección Heckman .....	82
4.4 Heckman Oaxaca y descomposición triple.....	87
4.5 Modelo espacial SLX (panel municipal): efectos fijos locales y de vecindad.....	92
4.5.1 Ajuste del modelo y validación espacial (2018 – 2024).....	96
4.5.2 Cartografía de rezago de formalidades y dinámica de clústeres.....	98
<b>Conclusiones.....</b>	<b>104</b>
<b>Bibliografía.....</b>	<b>108</b>
<b>Trabajos citados .....</b>	<b>108</b>
<b>Bibliografía complementaria .....</b>	112
<i>Anexo A. Descriptor de las variables .....</i>	115
<i>Anexo B. Manual de construcción de la base de datos ENOE .....</i>	116
<i>Anexo B2. Fase preliminar; conformación de la muestra. ....</i>	126
<i>Anexo C. Modelo Oaxaca Blinder, y Heckman .....</i>	131
<i>Anexo D. Modelo SLX implementación de econometría espacial .....</i>	132

## **CAPÍTULO I. Introducción y marco teórico**

### 1.1 Introducción al problema

La brecha de género constituye una de las manifestaciones más persistentes y complejas de la desigualdad estructural en los mercados laborales contemporáneos. En el caso de México, esta disparidad ha perdurado a pesar de décadas de esfuerzos institucionales dirigidos a promover la equidad de género. Lejos de ser una simple consecuencia de diferencias en capital humano entre hombres y mujeres, la brecha refleja una estructura multifactorial de barreras visibles y latentes, cuya interacción afecta de manera diferenciada las trayectorias laborales y las oportunidades económicas entre géneros. Esta investigación parte de la premisa de que la desigualdad salarial es tanto una expresión de inequidades históricas como una reproducción activa de condiciones socioeconómicas discriminatorias en el mercado de trabajo mexicano.

La evidencia empírica disponible para el país muestra que, incluso cuando hombres y mujeres presentan características observables similares, como años de escolaridad, experiencia laboral, ocupaciones o lugar de residencia, las diferencias en los ingresos persisten de manera sistemática. Este fenómeno pone en duda la suficiencia de las explicaciones basadas exclusivamente en el capital humano y obliga a incorporar en el análisis elementos estructurales no observables como la discriminación, los estereotipos de género, las normas culturales y la segmentación del mercado laboral.

Dentro de este contexto, el período 2019-2023 resulta particularmente relevante. Este intervalo temporal no solo permite observar la dinámica estructural de las brechas bajo condiciones económicas ordinarias, sino que también ofrece la oportunidad de evaluar los efectos disruptivos y diferenciados de la pandemia de COVID-19 en las condiciones laborales de mujeres y hombres. Este fenómeno ha sido documentado durante este período por otros autores: “Durante la pandemia de COVID-19, las mujeres mexicanas enfrentaron una carga desproporcionada de trabajo no remunerado y una mayor afectación en el empleo formal, lo que incrementó su vulnerabilidad en el mercado laboral, particularmente en sectores informales y de bajos ingresos” (Arceo-Gómez & Torres, 2021).

Asimismo, este estudio se enmarca en una creciente corriente analítica que reconoce la importancia de estimaciones econométricas rigurosas para la identificación y cuantificación de desigualdades de ingreso. En particular, se retoman modelos como la ecuación de Mincer, el método de corrección por sesgo de selección de Heckman y la descomposición Oaxaca-Blinder, herramientas que permiten no solo aislar el componente explicado por características observables, sino también identificar la porción no explicada de la brecha, atribuible potencialmente a mecanismos discriminatorios.

La persistencia de esta desigualdad tiene implicaciones profundas tanto a nivel individual como agregado. Desde una perspectiva microeconómica, las mujeres enfrentan penalizaciones que afectan su poder adquisitivo, sus incentivos al esfuerzo y su bienestar subjetivo. A nivel macroeconómico, la subutilización del talento femenino representa una pérdida significativa de eficiencia productiva y crecimiento potencial. Además, la desigualdad salarial contribuye a la feminización de la pobreza, afecta la autonomía económica y perpetúa ciclos de desventaja intergeneracional, particularmente en los hogares donde ellas son las principales proveedoras.

El análisis de la brecha salarial de género en México, por tanto, no puede limitarse a la comparación superficial de promedios. Es necesario un enfoque más robusto y estratificado que contemple las interacciones entre variables sociodemográficas, el entorno institucional, las estructuras regionales y las características del empleo formal e informal. La literatura reciente sugiere que las brechas tienden a ser más pronunciadas en contextos de baja regulación laboral, escasa cobertura de seguridad social y altos niveles de informalidad, condiciones todas presentes en gran parte del mercado laboral mexicano.

En este sentido el presente trabajo busca aportar evidencia empírica robusta sobre las causas y magnitudes de la desigualdad salarial entre hombres y mujeres en el país. El análisis se sustenta en microdatos provenientes de la Encuesta Nacional de Ocupación y Empleo (ENOE), los cuales permiten una exploración detallada de las diferencias salariales a través de distintos estratos poblacionales y temporales. A partir de ello, se pretende no solo documentar la existencia de brechas, sino también proponer interpretaciones que sirvan de base para el diseño de políticas públicas más efectivas, centradas en la equidad salarial y la justicia económica.

Finalmente, es fundamental resaltar que la brecha salarial de género no solo responde a mecanismos explícitos de discriminación, sino también a formas más sutiles de exclusión femenina en posiciones de liderazgo y la sobrecarga de trabajo doméstico no remunerado. Estos factores, al operar simultáneamente, configuran una estructura laboral que penaliza sistemáticamente al género femenino, aun cuando este cuenta con niveles crecientes de preparación académica y participación económica.

En suma, el abordaje de la brecha salarial desde un enfoque cuantitativo permite no solo delimitar el alcance del problema, sino también identificar sus determinantes específicos y estimar el peso relativo de los factores observables y no observables. Con ello, se pretende contribuir al desarrollo de un diagnóstico más preciso y orientado a la transformación de las estructuras que perpetúan la desigualdad salarial en México.

### 1.1.1 Justificación

El motivo de este trabajo radica en la necesidad de abordar una problemática que persiste en el ámbito social y económico en México, la brecha salarial que forma parte de uno de los factores más importantes de la desigualdad de género. Esta brecha es un reflejo de las diferencias estructurales y la discriminación como factores no explicativos que aún prevalecen y limitan el desarrollo económico de las mujeres.

En México, las mujeres suelen recibir salarios menores por realizar trabajos equivalentes en comparación con los hombres, incluso teniendo características similares (edad, experiencia o tipo de empleo, etc.), ya se ha comprobado que las mujeres deben enfrentarse a dificultades que afectan directamente a su desempeño laboral, como la capacidad de negociación de sueldos, maternidad o desempeño en áreas específicas de trabajo, además de estereotipos y sobrerrepresentación de áreas laborales de menor paga, al mismo tiempo se enfrentan a prácticas discriminatorias del mismo mercado laboral. Se ha observado que las mujeres tienen ocupaciones históricamente ligadas con roles de cuidado y hogar, como factores culturales que prevalecen. Estas diferencias perpetúan la subutilización del potencial productivo femenino. Esta disparidad no solo es una cuestión de justicia social, sino también un reflejo de una

estructura económica que castiga el trabajo femenino y genera barreras en el desarrollo personal y profesional, además, tiene consecuencias generacionales, afectando en mayor proporción a los hogares encabezados por mujeres.

La discriminación se presenta en múltiples escalas y en diferentes dimensiones, como la posición de liderazgo, la segregación ocupacional, y las condiciones laborales precarias, mismas que están ligadas a una estructura discriminatoria que se prolonga en los diferentes niveles de la sociedad, estas condiciones sociales afectan directamente la independencia económica, en las mujeres perpetúan ciclos de pobreza, ya que limitan su capacidad de consumo, limitando posibilidades como inversión en salud, educación, bienestar familiar, patrimonio e incluso ocio y entretenimiento. Son condiciones que no están aisladas, ya que forman parte inherente de la sociedad y de la población en la medida de sus características y en otra porción por aspectos sociales, culturales, tradicionalistas. Examinar este problema desde un enfoque econométrico nos permitirá cuantificar el peso de las variables socioeconómicas de la brecha salarial, de igual manera nuestro análisis permite evidenciar los vínculos relacionados con la discriminación.

Para comprender las principales repercusiones de la disparidad en los ámbitos geográficos, laboral, social y económico, resulta crucial analizar sus atributos y la dimensión de esta problemática. Un estudio econométrico e inferencial permitirá examinar los diversos estratos sociales, identificando donde se intensifica este fenómeno y bajo qué particularidades se presentan mayores obstáculos. Al considerar las características específicas de nuestra muestra, podremos contrastar el nivel de diferenciación en distintos entornos sociales, como las divergencias en el capital humano y el grado de trato desigual, o bien, identificar cualquier predisposición que resulte relevante para nuestro análisis.

Mediante el presente estudio, se busca discernir los principales obstáculos que confrontan las mujeres mexicanas para alcanzar una remuneración equiparable a la de los hombres. Los resultados obtenidos permitirán delimitar los grupos más vulnerables y los elementos que ejercen mayor influencia en esta disparidad. Esta información resulta fundamental para el diseño y promoción de estrategias políticas orientadas a mitigar la desigualdad en los ámbitos social, laboral y familiar. Asimismo, facilitará la focalización de esfuerzos en iniciativas de

capacitación, como la formulación de reformas regulatorias y mecanismos de supervisión en materia de equidad de género.

Nuestra investigación tiene como objetivo contribuir a la investigación de las desigualdades en materia de género, además de contribuir a las futuras investigaciones, a través de la generación de evidencia estadística y actualizada de México, tomando como marco temporal el período de pandemia, como un análisis particular de la recesión económica mundial a causa de esta, no solamente se busca contribuir al conocimiento académico mediante este trabajo, sino también la construcción de una agenda pública más justa, que puede servir en la formulación de políticas públicas más informadas y eficaces.

### 1.1.2 Preguntas de investigación

En el curso de la investigación se busca dar respuestas a interrogantes esenciales de las diferencias socioeconómicas y su relación con la brecha salarial, preguntas como: ¿En qué medida las características socioeconómicas (nivel educativo, edad, rama a la que pertenecen, etc.) explican las diferencias salariales entre hombres y mujeres?, de igual manera dimensionar el tamaño de la brecha mediante preguntas como: ¿Qué proporción de la brecha puede atribuirse a discriminación estructural no explicada por las variables observadas?, de lo cual posteriormente a través de la metodología pretendemos dar un modelo de explicación que determine parámetros de relación por características observables y no observables, de las cuales se despliegan preguntas como: ¿Cuál es la magnitud de la brecha salarial entre hombres y mujeres en los diferentes sectores e industrias? También dar respuesta y actualización a algunas problemáticas mediante el análisis estadístico a preguntas como: ¿Cuáles son las principales causas estructurales que contribuyen a la persistencia de estas brechas?

En segundo plano se busca responder a preguntas que resultan importantes en la desagregación de los resultados, como lo son: ¿En qué medida las características de capital humano explican la diferencia de salarios?, ¿Cómo afectan las normas culturales y los estereotipos de género a la distribución del trabajo y a las oportunidades de ascenso en el ámbito laboral?, ¿Cuál es el impacto de la brecha salarial de género en la economía general y el bienestar socioeconómico, en ambos géneros?

Estas preguntas de investigación son fundamentales para analizar la complejidad de la brecha salarial y laboral, mismas que se pueden utilizar en un análisis geográfico, o bajo contextos específicos. Al examinar las dimensiones socioeconómicas, estructurales, de capital humano, culturales y de impacto, esta investigación busca proporcionar una comprensión mejor integrada. Preguntas como cuáles son los principales factores estructurales o las características de mayor impacto en el salario pueden ser importantes en la formulación de políticas públicas. Algunas preguntas específicas como el impacto del capital humano nos pueden ayudar de forma pertinente para dotar de recursos a ciertas condiciones laborales en las que grupos sociales deben evitar, eliminar o incluso adoptar.

### 1.1.3 Objetivo

El objetivo de este trabajo de investigación es analizar el fenómeno de la persistente brecha salarial de género en el período de 2019 – 2023 en México, como punto fundamental de impacto de la pandemia, un momento previo a ella, durante la pandemia y cuando la mayor parte de la población ya estaba vacunada, a través de la Encuesta Nacional de Ocupación y Empleo (ENOE). Este fenómeno social es multifactorial, su complejidad no se debe limitar a la explicación frívola de la discriminación y los patrones culturales, sino profundizar en las características que la provocan.

Esta desigualdad no refleja la posición subordinada que históricamente han ocupado las mujeres en el mercado laboral, sino también las barreras sistemáticas que limitan su acceso a oportunidades de empleo equitativas y bien remuneradas. Aún con los esfuerzos normativos y la propiciación de mayor participación femenina en sectores productivos o con mayor representación femenina los datos muestran que las mujeres continúan recibiendo ingresos inferiores en comparación con la de los hombres, incluso bajo características similares, como educación, habilidades y experiencia. Entender las causas subyacentes será parte de nuestro estudio.

De igual manera el objetivo de esta investigación es analizar la evolución y los determinantes socioeconómicos de la brecha salarial de género en México durante el período 2019 – 2023, a

través de nuestro análisis en estadística podremos observar comportamientos, parámetros y tendencias de la población y de los grupos sociales, además, la aplicación de métodos econométricos nos permitirá distinguir factores explicables por características observables de la población muestral, como educación, edad, tipo de empleo, estado civil, tiempo de trabajo, etc., y aquellos factores no observables, vinculados con mecanismos de discriminación o sesgos sistemáticos del mercado laboral.

Este período de tiempo es importante señalar este período de tiempo, tiene cómo objetivo comprender el comportamiento dinámico a lo largo de este período y contrastarlo con años anteriores para observar cambios importantes generados por la pandemia COVID-19. En este período las condiciones laborales de la población se vieron afectadas, aún más las condiciones femeninas tanto en la pérdida de empleos en sectores feminizados, como el aumento de la carga de trabajo no remunerado en hogares, por ello es fundamental explorar cómo estos eventos influyen en las desigualdades salariales de hombres y mujeres.

Mediante la metodología de Oaxaca-Blinder podemos hacer un análisis econométrico para discernir la brecha salarial en sus componentes: sus características observables (edad, estado civil, nivel educativo, ocupación etc.) y el no explicado que refleja factores estructurales invisibles en los datos, misma que permite la evidencia de cómo las condiciones socioeconómicas pueden explicar las desigualdades económicas.

Como objetivos secundarios, nuestro trabajo de análisis ayuda a identificar principales características laborales que hacen vulnerables a ciertos grupos sociales, como la región geográfica o el nivel educativo, mismas que ayudan a la formulación de políticas públicas orientadas a mitigar la evolución de la brecha y dotar de facultades a los grupos más afectados por las condiciones laborales y económicas, además de aportar evidencia sólida para mejorar los marcos regulatorios en materia de salarios, fomentar la corresponsabilidad de trabajos, e incluso en futuras investigaciones puedan orientar sus investigaciones para el análisis de este fenómeno, este esfuerzo académico busca contribuir en la comprensión de un fenómeno social a través del estudio cuantitativo de características socioeconómicas para la promoción de políticas más justas.

#### 1.1.4 Hipótesis

La brecha salarial de género se entiende aquí como un fenómeno con muchas capas, que no se agota en las diferencias visibles de educación, experiencia o tipo de empleo entre personas ocupadas. Cuando se incorporan estas características productivas en los modelos y se revisan con calma los resultados, sigue apareciendo una parte del diferencial de ingresos que no se ajusta a las características individuales observadas. Ese “resto” de la brecha, que permanece incluso después de controlar por variables socioeconómicas, apunta a la presencia de mecanismos de discriminación más profundos dentro del mercado laboral mexicano. Con esta idea de fondo, la investigación plantea una hipótesis central: en México, durante el período de 2019 a 2023, las condiciones socioeconómicas observables permiten explicar solo una fracción de la brecha salarial entre hombres y mujeres, mientras que una proporción importante del diferencial persiste sin explicación estadística clara, lo que se interpreta como evidencia de la permanencia de prácticas de discriminación estructural en contra de las mujeres trabajadoras.

Esta hipótesis descansa en dos pilares teóricos principales. Por un lado, se retoman los aportes de la teoría del capital humano, que vincula las diferencias salariales con las dotaciones de educación, trayectoria laboral, tipo de ocupación y otras características asociadas a la productividad individual. Por otro lado, el análisis se complementa con las teorías de la discriminación de género, que señalan que, aun cuando hombres y mujeres presentan perfiles laborales muy similares, las decisiones de contratación, ascenso y fijación de salarios no son neutrales, sino que suelen inclinarse a favor de los hombres. En conjunto, estas perspectivas permiten interpretar que la parte de la brecha que permanece sin explicación no es simplemente “ruido” en los datos, sino la huella de relaciones de poder y sesgos de género que siguen presentes en el funcionamiento cotidiano del mercado de trabajo. Lo cual forma parte del componente no explicado de nuestro modelo teórico Oaxaca-Blinder, donde ya otros autores han encontrado estadísticamente significativa una porción representativa de desigualdad entre géneros. Estos supuestos ayudan a plantear hipótesis específicas como: “Las mujeres presentan una penalización salarial no explicada por su perfil productivo, lo cual refleja un mecanismo de discriminación estructural” o “la magnitud de la brecha salarial varía según el grupo poblacional, siendo más pronunciada en algunos sectores económicos, regiones geográficas o

niveles escolares”. La validación de estas hipótesis será posible bajo nuestro análisis cuantitativo y uso de técnicas econométricas con los datos provenientes de las encuestas nacionales y aplicación de estadísticos robustos que permitan descomponer la brecha salarial. Estas hipótesis tienen importancia bajo los resultados que el trabajo académico ponga en evidencia.

## 1.2 Revisión de literatura

### 1.2.1 Marco teórico general

La desigualdad de ingresos entre hombres y mujeres, conocida como la brecha salarial de género, ha sido objeto de diversas interpretaciones teóricas dentro de la economía del trabajo. Lejos de responder a una causa única o lineal, este fenómeno representa la interacción de múltiples dimensiones: factores de productividad individual, mecanismos institucionales, normas culturales persistentes y estructuras de poder que operan tanto dentro como fuera del mercado laboral. El presente apartado ofrece un recorrido analítico por los principales estudios que permiten integrar de forma integrada, las raíces económicas de esta desigualdad, centrándose especialmente en su expresión dentro del mercado mexicano.

En el análisis económico, los salarios son concebidos como precios relativos que reflejan, al menos teóricamente, la productividad marginal de los trabajadores. Sin embargo, este enfoque, asociado a la economía neoclásica, parte de un conjunto de supuestos que en la práctica no siempre se cumplen: mercados perfectamente competitivos, información simétrica, racionalidad plena y ausencia de barreras sociales (Blau & Khan, 2017). Estas condiciones laborales ideales invisibilizan otros factores que inciden en la determinación de los ingresos laborales, como la posición social, las restricciones institucionales o las dinámicas de negociación individual y colectiva.

Al asumir que el ingreso está determinado por las “dotaciones” del individuo (educación, experiencia, habilidades), se ignora que la valorización de estas cualidades no es neutral, sino que depende de estructuras históricas e institucionales. Así, las diferencias salariales por género no pueden entenderse solamente como un reflejo de “menor productividad” femenina, sino como el resultado de relaciones de poder que moldean la asignación de valor en el mercado laboral (Rodríguez Pérez, 2014).

La división sexual del trabajo constituye una de las formas más persistentes de segmentación ocupacional. Desde la perspectiva estructuralista, el mercado laboral no es homogéneo, sino que está organizado en compartimentos que ofrecen diferentes condiciones según el sector y el perfil del trabajador (Arceo-Gómez & Torres, 2021). Esta segmentación se expresa en una concentración de mujeres en sectores de baja remuneración como servicios personales, educativos o cuidados, y su subrepresentación en ocupaciones técnicas, directivas o institucionales.

Este fenómeno no es consecuencia de decisiones individuales libres, sino de una configuración institucional y cultural que canaliza a las mujeres hacia roles considerados “naturales” para su género. Desde edades tempranas, las trayectorias formativas y profesionales femeninas están condicionadas por estereotipos y expectativas que limitan sus opciones reales de inserción. Tal como lo menciona (Bonavida & Gasparini, 2024), la segmentación ocupacional actúa como una barrera de acceso a empleos mejor remunerados, consolidando una brecha salarial estructural.

En muchas organizaciones, la estructura jerárquica sigue mostrando una forma de discriminación menos visible pero igual de efectiva. Esta se expresa en la exclusión de las mujeres de los cargos de mayor nivel, lo que limita su participación en la toma de decisiones y frena su desarrollo profesional. Este fenómeno, identificado como techo de cristal, obstaculiza su acceso a puestos de liderazgo o alta dirección, incluso cuando poseen una preparación y una experiencia comparables o superiores a las de sus colegas hombres (Hedija, 2023).

Uno de los elementos menos visibles en los estudios económicos tradicionales es la dimensión temporal del trabajo y su relación con la desigualdad de género. En sociedades como la mexicana, donde persisten modelos familiares tradicionales, las mujeres siguen asumiendo una parte desproporcionada del trabajo doméstico y de cuidado (Babcock et al., 2003). Esta doble jornada impacta directamente en su disponibilidad para insertarse en empleos de tiempo completo, asumir responsabilidades extendidas o acceder a oportunidades de desarrollo profesional. Esta restricción temporal también frena la acumulación de capital humano específico, por ejemplo, la experiencia laboral continua y la formación en el puesto, y en consecuencia limita la posibilidad de acceder a mejores ingresos a lo largo de la trayectoria laboral (Killingsworth, 1983).

Desde la economía del hogar, se ha señalado que esta distribución desigual del tiempo genera una división funcional entre quienes están disponibles para el mercado y quienes asumen los costos sociales de la reproducción cotidiana (Oaxaca & Ransom, 1994).

Más allá de las características individuales y las normas sociales, el marco institucional en el que se desenvuelve el mercado laboral tiene un papel decisivo en la configuración de la desigualdad salarial. La legislación, los convenios colectivos, los sistemas de seguridad social y los mecanismos de fiscalización definen los límites y posibilidades para la igualdad de oportunidades (Salce Díaz, 2021).

En contextos donde las instituciones laborales son débiles, como es el caso en buena parte del mercado mexicano, las normas de equidad salarial suelen quedar en el plano declarativo. La falta de transparencia en los procesos de contratación, promoción y remuneración permite que las decisiones empresariales se vean influenciadas por sesgos de género sin mayores consecuencias (Rodríguez Pérez & Limas Hernández, 2017).

La ausencia de políticas efectivas para conciliar vida laboral y familiar, la escasa cobertura de servicios de cuidados públicos y la baja participación femenina en sindicatos o instancias de negociación colectiva, también contribuyen a reforzar estas dinámicas. En consecuencia, las mujeres enfrentan condiciones de desventaja no solo en el acceso al trabajo, sino también en su permanencia y progresión dentro del mismo (Scotti & Manzano, 2015).

La desigualdad salarial por género no puede analizarse únicamente desde el promedio de ingresos, ya que esta presenta una estructura heterogénea a lo largo de toda la distribución salarial. Desde una perspectiva cuantílica, se revela que las disparidades entre hombres y mujeres se intensifican en distintos tramos de dicha distribución, haciendo evidente que la discriminación no es uniforme. A medida que algunas mujeres logran ascender en la jerarquía ocupacional, otras enfrentan barreras persistentes que las confinan a los segmentos más bajos del ingreso, consolidando fenómenos como el “piso pegajoso”, mientras que otras, incluso con alta cualificación, son excluidas de los puestos mejor remunerados por un “techo de cristal” que restringe su movilidad ascendente. Estos patrones no son meramente anecdóticos, sino que reflejan una segmentación estructural del mercado laboral que penaliza de forma diferenciada la participación femenina según el nivel salarial, configurando una estratificación que sólo

puede ser plenamente comprendida mediante técnicas econométricas como la regresión cuantílica, que permite capturar la dinámica dispereja de las brechas en los distintos percentiles de la distribución (Yusuf et al., 2023). Estos datos resultan cruciales para revelar los mecanismos que refuerzan las desigualdades tanto en la base como en la cúspide del sistema ocupacional.

Otros autores han estudiado la diferencia salarial de género a través de la dimensión interseccional de vulnerabilidad acumulada, esta desigualdad responde a una compleja interacción de factores estructurales que agravan o atenúan su impacto según las características individuales y contextuales de cada mujer. Variables como el nivel educativo, la ubicación geográfica, la pertenencia étnica, la edad o el entorno rural-urbano, configuran escenarios diferenciados de exclusión laboral, siendo especialmente pronunciadas las desventajas entre mujeres indígenas, residentes rurales o aquellas en etapas extremas del ciclo de vida, como la juventud o la vejez (Luz Tovar & Pineda Salazar, 2023). En este sentido, el enfoque interseccional ofrece un marco analítico imprescindible entre sí, generando realidades laborales más precarias para quienes acumulan condiciones de vulnerabilidad. Así, por ejemplo, una mujer con escasa escolaridad que habita en una región con alta informalidad no solo enfrenta barreras educativas o territoriales, sino una combinación sinérgica de ambas que agrava su exclusión del mercado laboral (Albuja & Enríquez, 2018). Ante ello, resulta indispensable que cualquier diagnóstico econométrico y formulación de políticas públicas considere esta diversidad de trayectorias, a fin de que las estrategias de intervención no partan de un supuesto homogéneo, sino que reconozcan la pluralidad de desafíos que enfrentan las mujeres en el ámbito salarial.

Desde la macroeconomía, la brecha salarial no es solo un asunto de justicia social. También implica costos de eficiencia económica. Cuando el mercado penaliza de forma sistemática el trabajo femenino, se desaprovecha una parte importante del capital humano, se distorsionan los incentivos de asignación de recursos y se reduce la productividad agregada (Buser et al., 2018).

La exclusión de las mujeres de sectores estratégicos, la desvalorización de sus aportes y la limitación de sus oportunidades laborales no solo reducen su bienestar individual, sino que afectan el crecimiento potencial de la economía. Estudios recientes han mostrado que las

economías con mayores niveles de equidad de género tienden a registrar mayores tasas de crecimiento, innovación y desarrollo humano (Otsu & Tanaka, 2022).

En este sentido, reducir la brecha salarial no es únicamente una demanda feminista o un imperativo moral, sino también una condición para el desarrollo sostenible y la eficiencia sistemática. La igualdad de género en el mercado de trabajo debe ser entendida como una inversión social económica, no como una concesión.

Parte del trabajo también aborda la relación territorial con las brechas, en el estudio espacial, entre ambos sexos, buscando comprender otro panorama de la trayectoria del ingreso individual, autores como Rosales y Mendoza (2023), han estudiado el caso del sector turístico, y las mujeres en industria, dentro de sus hallazgos esta existe una significativa diferencia en una remuneración relativamente menor. De igual manera Castro Romero y Barroso (2025), en el sector turístico, específicamente centran su análisis en la industria hotelera, se observa un acceso femenino a cargos gerenciales “limitado”, mismo que apunta a un mecanismo de discriminación estructural, esto puede ser un efecto de techo de cristal.

Análisis espaciales como este y el de (Smith & R., 2013), cuestionan la relación espacial como problema a la desigualdad salarial de género, ya que puede amplificar brechas persistentes, tomando en cuenta que el lugar donde se vive penaliza los retornos, una observación importante en relación geográfica, que enraíza patrones espaciales desiguales.

Evidencias empíricas en México han analizado que la distribución territorial, puede concentrar precariedad laboral lo cual se ve reflejado mediante patrones de autocorrelación espacial significativa, áreas geográficas concentran desigualdades pronunciadas (Andres-Rosales et al., 2019), en complemento el trabajo de Castro et al., (2021) indica que no se ha observado convergencia regional en la brecha salarial de género entre 2005 y 2020. En lugar de ello, se registra una tendencia a la persistencia territorial de las desigualdades, con estructuras locales que tienden a reforzarse en el tiempo.

En términos metodológicos, la literatura reciente ha puesto especial énfasis en el análisis espacial para estudiar desigualdades. Se utilizan métricas de autocorrelación —tanto globales como locales, entre ellas el índice de Moran y los indicadores LISA—, así como modelos de regresión espacial (SAR, SEM, SDM) y esquemas de panel con efectos territoriales (Elhorst,

2003; Anselin, 2022). Estas herramientas hacen posible no sólo capturar patrones de concentración regional, sino también evaluar las interdependencias entre áreas vecinas, lo cual resulta esencial para comprender cómo se distribuyen los procesos de desigualdad en distintos contextos.

En el plano comparativo (Matuszewka-Janica, 2025) muestra que la brecha salarial de género sin ajustar en Europa carece de una autocorrelación espacial clara a nivel nacional; sin embargo, a escalas regionales surgen evidencias de dependencia que apuntan a la influencia de factores estructurales como las trayectorias migratorias, la especialización productiva y las disparidades en infraestructura. En el caso mexicano, estudios como los de (Hernández Ballesteros & Almonte, 2024; Romero & Prudencio, 2022) han cartografiado desigualdades municipales y estatales mediante estadística espacial descriptiva, identificando asociaciones entre la estructura económica local y la intensidad de la brecha. No obstante, la ausencia de un tratamiento econométrico formal limita la posibilidad de inferir causalidad o de distinguir efectos vecinales, lo que refuerza la pertinencia de introducir modelos como el SLX en análisis posteriores.

De manera complementaria, la literatura también ha incorporado enfoques novedosos que amplían las posibilidades del análisis. (Lombardia et al., 2022) proponen una descomposición sustentada en teoría de asignación y búsqueda, la cual permite incluir heterogeneidad no observada y simulaciones contrafactuales, abriendo paso a la identificación de interacciones no lineales entre género, sector y ocupación. De igual forma, Lummertz y colaboradores han explorado la aplicación de redes neuronales en modelos de crecimiento económico espacial, demostrando que técnicas de inteligencia artificial pueden capturar estructuras territoriales de gran complejidad. Estas innovaciones fortalecen la capacidad de diagnóstico sobre las desigualdades de género, al integrar simultáneamente dimensiones individuales, sectoriales y espaciales, lo que representa un paso importante para generar evidencia robusta que sustente políticas públicas con enfoque territorial e interseccional.

### 1.2.2 Segregación ocupacional y brecha de género

Uno de los factores estructurales más persistentes que profundizan la brecha salarial entre hombres y mujeres es la segregación ocupacional, entendida como la distribución desigual de los sexos en los diferentes tipos de empleos y sectores de actividad. Esta segmentación

responde tanto a patrones históricos como a construcciones socioculturales que asocian determinadas actividades laborales con estereotipos de género. En consecuencia, las mujeres tienden a concentrarse en ocupaciones que, aunque intensivas en trabajo, son sistemáticamente menos valoradas y remuneradas por el mercado.

En el contexto mexicano, esta segregación se expresa de múltiples formas: horizontalmente, cuando las mujeres y hombres se ubican en diferentes sectores productivos; y verticalmente, cuando existen barreras para el ascenso femenino a puestos de mayor jerarquía y responsabilidades. A pesar de los avances en cobertura educativa y participación laboral, las mujeres continúan sobrerrepresentadas en sectores de servicios, educación, salud y cuidado, áreas asociadas históricamente al trabajo doméstico y reproductivo, y por tanto con bajos niveles salariales.

Este fenómeno no es exclusivo de México, pero encuentra particular agudeza en su mercado laboral debido a la interacción entre la cultura patriarcal, la informalidad estructural y las débiles políticas de conciliación entre vida laboral y familiar. Estudios como los de (Albuja & Enríquez, 2018) evidencian que, incluso dentro del sector formal, la asignación ocupacional mantiene sesgos de género. El llamado “techo de cristal”, barrera invisible que impide a las mujeres ascender en la jerarquía laboral pese a contar con méritos similares a sus pares varones, es una manifestación directa de esta lógica discriminatoria que incide en la estructura ocupacional.

Desde una perspectiva cuantitativa, la segregación ocupacional se convierte en un mecanismo clave para explicar la porción explicada de la brecha salarial de género. Sin embargo, aun controlando estadísticamente por tipo de empleo, nivel educativo y experiencia, persiste una fracción significativa de la brecha que no puede justificarse por diferencias observables, lo que apunta a la existencia de discriminación estructural. Tal como señala (Fortin, 2011), el análisis cuantílico de la brecha muestra que el efecto del género es más pronunciado en los extremos de la distribución, especialmente en los percentiles bajos (piso pegajoso) y altos (techo de cristal), evidenciando una penalización adicional en estos grupos.

La literatura ha abordado el papel de la autoselección y de las expectativas sociales en la asignación ocupacional. En contextos de socialización diferenciada, las mujeres no solo enfrentan barreras externas, sino que también arrastran restricciones internas que moldean sus

decisiones educativas y laborales. Como señalan Blau y Kahn (2017), estas construcciones sociales influyen en la negociación salarial, en la elección de carrera y en la proyección del futuro laboral, y reducen las oportunidades de acceso a sectores masculinizados o de mayor ingreso.

La segregación, por tanto, no debe entenderse como una elección libre. Más bien es el resultado de un conjunto de restricciones que interactúan entre sí, como las institucionales, culturales y económicas, que limitan las trayectorias femeninas y refuerzan estructuras desiguales. Analizarla permite visibilizar cómo el mercado laboral reproduce jerarquías de género en el plano ocupacional. Esto resulta clave para diseñar políticas orientadas a la equidad, entre ellas la valoración del trabajo de cuidado, la aplicación de cuotas en cargos directivos y la promoción de mujeres en sectores estratégicos.

En conjunto, la segregación ocupacional es un punto crítico para entender la brecha salarial de género. Su permanencia impide que el avance educativo de las mujeres se traduzca por completo en mejores resultados laborales. Se genera un desajuste estructural entre capacidades y recompensas. Superar esta barrera exige cambiar no solo las reglas del mercado de trabajo, sino también los imaginarios que sostienen las divisiones del trabajo.

### 1.2.3 Brecha salarial y estructura del mercado laboral mexicano

La persistente brecha salarial entre mujeres y hombres en México refleja una estructura laboral fuertemente segmentada, con alta informalidad, rigideces institucionales y desigualdades regionales y sectoriales marcadas. En este marco, la distribución del ingreso no depende solo de diferencias observables en capital humano. También interviene una arquitectura económica que genera y refuerza mecanismos de exclusión, muchas veces poco visibles, que afectan de manera diferenciada las trayectorias laborales de las mujeres.

Una primera característica relevante del mercado laboral mexicano es su elevada dualidad entre formalidad e informalidad. Esta última, entendida como aquella condición laboral que opera al margen de la regulación estatal (sin prestaciones, sin contrato, sin seguridad social), afecta de manera desproporcionada a las mujeres. De acuerdo con otras cifras de la ENOE, más del 55%

de las mujeres ocupadas entre 2019 y 2023 estuvieron empleadas en condiciones de informalidad, situación que contrasta con el 49% de los hombres. Esta disparidad no es casual ni coyuntural; obedece a dinámicas estructurales que colocan a las mujeres en sectores económicos de baja productividad, menor regulación institucional y escasa movilidad ascendente.

En efecto, sectores como el comercio al por menor, los servicios personales y el trabajo doméstico concentran una proporción significativa del empleo femenino, y se caracterizan por salarios bajos, alta rotación, condiciones precarias y mínima posibilidad de promoción. Según (Rodríguez Pérez R. E., 2019) esta segmentación laboral femenina tiene efectos directos sobre su nivel de ingreso, aun cuando se controlen características como educación, edad o experiencia. Es decir, existe una penalización estructural derivada del tipo de inserción sectorial, que acentúa la desigualdad salarial de género en contextos donde el Estado no regula de forma efectiva el funcionamiento del mercado de trabajo.

A esto se suma una segunda dimensión estructural: la rigidez institucional y la escasa aplicación de políticas laborales con enfoque de género. En México, a pesar de que la ley Federal del Trabajo establece la igualdad de remuneraciones por trabajos de igual valor, su cumplimiento es parcial y difícilmente fiscalizado. No existen mecanismos institucionales robustos que garanticen auditorías de equidad salarial en las empresas ni que penalicen de forma sistémica la discriminación remunerativa. Esta débil institucionalidad permite que persistan prácticas de desigualdad, tanto en los salarios de ingreso como en los ascensos, sin consecuencias jurídicas para los empleadores.

La literatura económica ha señalado que este tipo de entornos laborales carentes de mecanismos redistributivos y de rendición de cuentas, refuerzan el papel de la discriminación como factor determinante de la brecha salarial. En línea con (Becker, 1971) y su teoría de la “discriminación por gustos”, algunos empleadores internalizan preferencias personales que penalizan a las mujeres, ya sea por su rol potencial como madres, por prejuicios asociados a su “menor productividad” o simplemente por prácticas de contratación sesgadas. Cuando no existe vigilancia institucional, estas decisiones se reproducen y consolidan como parte del funcionamiento regular del mercado.

Un tercer elemento clave es la falta de políticas de conciliación entre la vida laboral y familiar. Las mujeres mexicanas siguen asumiendo una proporción desproporcionada del trabajo doméstico y de cuidado no remunerado, lo cual limita su disponibilidad de tiempo, reduce su flexibilidad laboral y restringe sus posibilidades de insertarse en empleos de tiempo completo o de alta exigencia horaria. Las sobrecargas domésticas, la maternidad y la capacidad de negociación de sueldos, no son neutras en términos económicos: genera trayectorias laborales más discontinuas, menor experiencia acumulada y, por ende, menores ingresos.

Además, en el período analizado 2019 – 2023, la pandemia por COVID-19 acentuó estas desigualdades. Según (Arceo-Gómez & Torres, 2021) las mujeres sufrieron mayores tasas de pérdida de empleo, sobre todo en sectores informales, además de que enfrentaron un incremento significativo en la carga de cuidados al interior del hogar. Estas dinámicas produjeron un efecto regresivo sobre la participación laboral femenina, evidenciando que la estructura del mercado mexicano no es resiliente ni equitativa frente a choques externos.

Desde una perspectiva metodológica, estas desigualdades estructurales se capturan parcialmente mediante el modelo Oaxaca-Blinder, el cual permite descomponer la brecha salarial entre una parte explicada -asociada a diferencias en dotaciones de capital humano- y una parte no explicada que puede interpretarse como indicio de discriminación. En el caso de México, estudios como el de (Rodríguez Pérez R. E., 2014) han estimado que más del 85% de la brecha salarial no puede ser atribuida a diferencias observables, lo que sugiere la existencia de mecanismos estructurales de exclusión en el funcionamiento del mercado laboral.

Un elemento por considerar es la segmentación regional y geográfica del mercado laboral mexicano. Las desigualdades en los ingresos no son homogéneas a nivel nacional; existen profundas brechas entre zonas urbanas y rurales, así como entre regiones del norte y del sur del país. De acuerdo con (Luz Tovar & Pineda Salazar, 2023) en el sur de México más del 56% de la diferencia salarial entre hombres y mujeres indígenas no puede ser explicada por variables observables, lo que sugiere una fuerte discriminación en contextos donde la informalidad, la pobreza estructural y la debilidad institucional son más intensas. Este hallazgo es relevante para el análisis de brechas salariales en el país, pues demuestra que las estructuras del mercado

laboral no solo penalizan a las mujeres en general, sino con mayor intensidad a aquellas pertenecientes a grupos vulnerables o minoritarios.

A la par de esta evidencia, también debe señalarse que la estructura del mercado laboral opera con lógicas de penalización interseccional. Es decir, no todas las mujeres experimentan la misma brecha salarial. Aquellas con menor nivel educativo, residentes en zonas rurales o indígenas, y con responsabilidades familiares, enfrentan condiciones laborales significativamente más precarias que sus contrapartes con mayor escolaridad o que residen en grandes zonas urbanas. Esto refuerza la necesidad de que los análisis de la brecha salarial incorporen una perspectiva multidimensional y no se limiten a promedios agregados que ocultan desigualdades más profundas.

### 1.3 Marco metodológico general

El presente estudio se inscribe dentro del paradigma cuantitativo y posee un enfoque explicativo e inferencial, al buscar identificar y descomponer los determinantes de la brecha salarial de género en México durante el período 2019-2023. La investigación es de tipo no experimental, ya que no se manipulan deliberadamente las variables, sino que se observa su comportamiento en condiciones naturales, a partir de datos secundarios recopilados por INEGI. Asimismo, tiene un carácter transversal repetido en tanto que trabaja con múltiples cortes temporales independientes (año 2019, 2021, y 2023), permitiendo así identificar patrones y transformaciones interanuales relevantes especialmente frente al impacto de la pandemia de COVID-19.

Desde una perspectiva metodológica, se emplean modelos econométricos multivariados que permiten cuantificar la brecha salarial entre hombres y mujeres, distinguiendo entre la parte explicada por diferencias observables y la no explicada, presumiblemente atribuible a discriminación estructural. Las técnicas aplicadas incluyen la regresión mincerina, el modelo de corrección de sesgo de selección de Heckman, la descomposición Oaxaca-Blinder en su versión doble y triple.

La fuente de datos

La información proviene de la ENOE, en sus versiones correspondientes a los años 2019, 2021 y 2023. Cada año se compone de cuatro trimestres y cinco cuestionarios por trimestre: vivienda, hogar, sociodemográfico, ocupación y empleo (1 y 2). Estos fueron integrados utilizando procedimientos de combinación horizontal (*merge*) y vertical (*append*) mediante el software estadístico Stata 17. La muestra se restringió a personas de 15 años o más, que contestaron la totalidad del cuestionario. El manual de construcción de datos, conformación de variables y ajuste de modelos se encuentra en el apartado de anexos.

Para garantizar la representatividad estadística de los resultados, se incorporaron los factores de expansión poblacional proporcionados por INEGI, los cuales permiten extrapolar las estimaciones al universo poblacional. Esto asegura que los resultados reflejen fielmente la estructura sociodemográfica y laboral del país (Instituto Nacional de Estadística y Geografía, s.f.).

#### *Modelo Mincer*

La estimación inicial de la brecha salarial parte del modelo (Mincer, 1974), el cual permite aproximar los ingresos laborales en función del capital humano individual. Su especificación funcional básica es:

$$\ln(w_i) = X'_i\beta + \lambda_i\rho\sigma_u + \varepsilon_i$$

Donde:

- $\ln(w_i)$ : Logaritmo del salario (variable dependiente |salario).
- $X_i$ : Vector de variables explicativas:  
 $X_i = [\text{ur}, \text{grupo\_edad}, \text{grado\_esc}, \text{años\_esc}, \text{e\_civil}, \text{seg\_soc}, \text{rama}, \text{ing7c}, \text{dur\_est}, \text{serv\_med}]$
- $\beta$ : Coeficiente del Mills Ratio derivado de la etapa de selección.
- $\rho\sigma_u$ : Coeficiente que captura la correlación entre los errores de las dos ecuaciones.
- $\varepsilon_i$ : término de error, con  $E[\varepsilon_i|S_i = 1] = 0$

*Corrección por sesgo de selección: Modelo de Heckman*

Uno de los principales problemas en el análisis laboral es que los salarios solo son observados para personas ocupadas. Esto genera un sesgo de selección, ya que la decisión de participar en el mercado laboral no es aleatoria. Para corregir esta distorsión, se aplica el modelo de corrección de Heckman (1994), el cual consta de dos etapas:

*Etapas 1 (selección):*

$$S_i^* = Z_i\gamma + u_i \text{ donde } S_i = \begin{cases} 1 & \text{si } S_i^* > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Donde  $Z_i$  incluye variables que afectan la participación (estado civil, hijos, educación).

*Etapas 2 (ecuación de salarios)*

$$\ln(\text{Salario}_i) = X_i\beta + \rho\lambda_i + \varepsilon_i$$

Donde  $\lambda_i$  es el inverso del Mills ratio, calculado a partir de los residuos del modelo Probit, y  $\rho$  mide la correlación entre los errores de las dos ecuaciones. Esta corrección permite obtener estimaciones insesgadas de los determinantes del salario.

*Descomposición Oaxaca Blinder*

La metodología Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) permite descomponer la brecha salarial promedio entre hombres y mujeres en componentes explicados y no explicados. En su versión básica:

$$\Delta\bar{Y} = (\bar{X}_H - \bar{X}_M)\hat{\beta}_M + \bar{X}_M(\hat{\beta}_H - \hat{\beta}_M)$$

Donde  $(\bar{X}_H - \bar{X}_M)\hat{\beta}_M$  representa la parte explicada por diferencias en dotaciones observables, mientras que  $\bar{X}_M(\hat{\beta}_H - \hat{\beta}_M)$  refleja la parte no explicada atribuible a retornos<sup>1</sup> diferenciados o discriminación.

Para un análisis más exhaustivo, se utiliza la descomposición triple propuesta por (Jann, 2008), la cual introduce un término de interacción:

---

<sup>1</sup> Retornos hace referencia a cuanto “paga” el mercado laboral por una característica (educación, experiencia/antigüedad, foralidad, sector etc.) en términos porcentuales del salario, manteniendo todo lo demás constante.

$$\Delta\bar{Y} = (X_H - X_M)\beta_M + X_M(\beta_H + \beta_M) + (X_H - X_M)(\beta_H - \beta_M)$$

Este enfoque permite aislar con mayor precisión los factores que contribuyen a la desigualdad salarial.

*Ajuste de la descomposición por selección: Oaxaca-Blinder*

Con el fin de integrar la corrección por selección a la descomposición salarial, se incorpora la diferencia en los Mills ratio entre grupos. La expresión extendida queda:

$$\Delta \ln(\text{Salario}) = (X_H - X_M)\hat{\beta}_M + X_M(\hat{\beta}_H + \hat{\beta}_M) + (\lambda_H - \lambda_M)\rho\sigma_\varepsilon$$

Este término estima la parte de la brecha derivada de las diferencias en la participación laboral, permitiendo una aproximación más realista al incorporar las restricciones estructurales que enfrentan las mujeres en el acceso al empleo.

*Modelo SLX (Spatial Lag of X)*

Este modelo permite obtener la influencia de las condiciones de los municipios vecinos sobre la desigualdad salarial local sin asumir dependencia espacial ni en la variable dependiente ni en los términos de error, a través de los efectos fijos de un panel de datos, 2018 – 2024, contrastando con especificaciones de efectos aleatorios, al que se incorpora una extensión de rezagos espaciales de las variables explicativas.

Donde nuestra variable dependiente es la brecha salarial por género, definida como la diferencia del ingreso promedio anual entre hombres y mujeres del cual se ha sometido a una previa intervención, para mitigar la influencia de outliers se aplicó la metodología de winsorización al 1% y 99%, del cual se trabajó con el logaritmo del salario winsorizado, facilitando la lectura del salario a través de semielasticidades, a partir de este punto, se han construido brechas que más adelante se denominan “gaps”, mismas que son producto de la diferencia de las características de hombres y mujeres. De estas “gaps” se han de estimar los efectos espaciales, se calcularon los rezagos espaciales de las variables explicativas mediante una matriz de contigüidad de primer orden (criterio *rook*), construida anualmente con base en los polígonos municipales disponibles. Para cada año, se generó una matriz de pesos espaciales  $W_t$  binaria y estandarizada por filas, que permite calcular el valor promedio ponderado de cada variable en los municipios

vecinos inmediatos (Elhorst, 2003). Formalmente, para cada variable explicativa X, el rezago espacial se define como:

$$WX_{it} = \sum_{i \neq j} w_{ij} X_{jt}$$

Donde

- $WX_{it}$ : representa el rezago espacial de la variable X en el municipio  $i$  en el año  $t$ ;  $X_{it}$  el valor de la misma variable en el municipio  $j$ ;
- $w_{ij}$  son los elementos de la matriz de pesos espaciales  $W$  definida con base a contigüidad geográfica.

Este procedimiento se aplica a cada uno de los años de manera independiente, asegurando la consistencia entre los identificadores geográficos de los datos y la geometría municipal, se corrigieron inconsistencias topológicas mediante un umbral de tolerancia (*snap distance*) y se empleó la opción `zero.policy = TRUE` para evitar errores en municipios sin vecinos registrados.

Posteriormente acompañado de un análisis de correlaciones y selección de criterio de relevancia sustancial, se definió el conjunto de variables explicativas por “gaps” brechas. Si bien algunas de estas variables presentan varianza limitada, su inclusión se justificó por su importancia en la estructura espacial del mercado laboral. Para cada una de ellas, se calcularon los rezagos espaciales anuales mediante la matriz de contigüidad geográfica, generando así el conjunto completo de predictores que capturan tanto efectos locales ( $\beta$ ) como de contexto vecinal ( $\theta$ ).

Una vez consolidada la base panel y construidos los rezagos espaciales para cada variable explicativa, se estimó un modelo econométrico tipo SLX (Spatial Lag of X) (Anselin, 2022; Elhorst, 2003). El modelo considera como variable dependiente la brecha logarítmica de ingresos laborales por género (*brecha\_log*) y como regresores tanto las variables estructurales locales como sus rezagos espaciales. La formulación general del modelo puede expresarse como:

$$y_{it} = \alpha_i + \lambda_i + X_{it}\beta + WX_{it}\theta + \varepsilon_{it}$$

Donde:

- $y_{it}$  : representa la brecha salarial por género en el municipio  $i$  y año  $t$ ,
- $a_i$  : Son efectos fijos municipales
- $\lambda_i$  : efectos temporales
- $X_{it}$  el vector de variables explicativas
- $WX_{it}$  : sus rezagos espaciales
- $\beta$  el coeficiente mide el efecto de la variable explicativa en el propio municipio sobre su propia brecha salarial.
- $\theta$  : es el efecto que tienen las condiciones promedio de los municipios vecinos sobre la brecha salarial local.

## CAPÍTULO II. Estadística descriptiva y caracterización de la muestra

### 2.1 Panorama estadístico

Este capítulo presenta información estadística importante previa a los modelos econométricos como panorama general de la muestra de las bases de datos de la ENOE, aquí se abordan las principales características sociodemográficas y laborales de la población mexicana mayor a 14 años, en tres cortes de tiempo 2019, 2021 y 2023. Estos estadísticos muestran una primera evidencia de la desigualdad en la población.

Cuadro 2.1 Estadísticas descriptivas. Población mayor a 14 años (2019, 2021 y 2023).

	2019		2021		2023	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
* N	409,983	441,434	690,687	744,145	749,242	798,418
**Salario	2,821.60	1,323.49	3,149.06	1,527.45	3,783.02	1,840.56
Edad	31.25	33.06	32.45	34.32	33.26	34.99
Casados	0.42	0.40	0.45	0.41	0.46	0.42
No sabe leer	0.06	0.07	0.06	0.07	0.06	0.06
Estudios						
< Primaria	0.16	0.15	0.14	0.13	0.13	0.13
Primaria	0.27	0.27	0.27	0.27	0.25	0.26
Secundaria	0.24	0.26	0.24	0.24	0.24	0.24
Preparatoria	0.17	0.15	0.18	0.16	0.19	0.17
Universidad	0.16	0.18	0.18	0.20	0.20	0.20
Grupo de edad						
≤ 14 años	0.27	0.23	0.25	0.22	0.24	0.21
14 a 24 años	0.18	0.18	0.18	0.16	0.18	0.16
24 a 34 años	0.14	0.14	0.14	0.15	0.14	0.14
34 a 44 años	0.13	0.14	0.13	0.14	0.13	0.14
44 a 54 años	0.12	0.12	0.12	0.14	0.12	0.13
54 a 64 años	0.08	0.10	0.09	0.10	0.09	0.10
≥ a 65 años	0.08	0.09	0.09	0.10	0.10	0.10
PEA	0.57	0.36	0.57	0.35	0.59	0.38

Nota: \*Contempla a la población que concluyó satisfactoriamente la entrevista. \*\* El salario representa el promedio de todas las observaciones.

Fuente: Creación propia con datos tomados de la ENOE (INEGI) 2019, 2021 y 2022.

La población observada para este estudio corresponde a quienes completaron satisfactoriamente la entrevista. Un primer elemento notable es que en los tres años analizados las cantidades de mujeres encuestadas superan de forma sistemática a la de los hombres. Por ejemplo, en 2023 se reportaron 798,419 mujeres frente a 749,242 hombres, lo cual representa una diferencia de aproximadamente 49 mil observaciones adicionales. Este comportamiento podría atribuirse a varios factores como una esperanza de vida mayor en las mujeres, en segundo lugar, la mayor disposición o permanencia en el hogar, lo cual aumenta la probabilidad de ser encuestadas por los levantadores de campo.

Una de las brechas más visibles y estructuralmente persistentes en el mercado laboral mexicano es la diferencia en los ingresos entre hombres y mujeres. En todos los años analizados, los hombres reportan ingresos laborales promedio muy por encima de los de las mujeres. En todos los años analizados, los hombres reportan ingresos laborales promedio muy superiores a los de las mujeres. En 2019, ellos promediaron \$2,821.60 y ellas \$1,323.29; dicho de otro modo, el ingreso femenino representó apenas el 46.9% del masculino. La relación mejora de forma ligera en los cortes siguientes, con 48.5% en 2021 y 48.6% en 2023, aunque la brecha sigue siendo amplia. Al no controlar por variables como horas trabajadas, ocupación o formalidad, esta diferencia refleja la brecha salarial bruta, que posteriormente será aislada de efectos observables y no observables.

La edad promedio también presenta una diferencia relevante y poco evidente a simple vista, En todos los años las mujeres entrevistadas con ligeramente mayores que los hombres en 2023 por ejemplo, la edad promedio femenina es de 34.99 años, mientras que la masculina es de 33.26, una diferencia de casi dos años completos puede estar vinculada con una esperanza de vida mayor en las mujeres, además de patrones de inserción y salida del mercado laboral.

La proporción de personas casadas también muestra diferencias sutiles pero significativas. En los tres años, una mayor proporción de hombres declara estar casado (46% en 2023) frente al 42% en mujeres. Esto puede estar asociado a el hecho de que muchas mujeres que crían solas no se declaran casadas aun teniendo hijos o pareja estable y una mayor proporción de jefatura femenina en hogares monoparentales, especialmente tras procesos de separación o viudez, más frecuentes en población femenina adulta.

El desglose por nivel de estudios muestra un dato clave: a partir de bachillerato y, sobre todo, en la universidad, las mujeres aparecen ligeramente más concentradas en los tramos altos que los hombres. En 2023, el 20% de ambos sexos reporta educación universitaria, lo que sugiere una paridad en el papel. Sin embargo, esa igualdad educativa no se traduce en salarios equivalentes. La literatura documenta una penalización en el retorno a la educación para las mujeres (Fortin, 2011). En consecuencia, aún con el mismo nivel educativo, su ingreso laboral suele ser menor.

Por lo tanto, en los niveles más bajos de escolaridad, la diferencia por sexo se reduce, lo cual indica que el rezago educativo en México sigue siendo una problemática transversal. En términos estructurales, la presencia de mujeres con alta escolaridad y bajos ingresos constituye una fuerte evidencia de segmentación vertical en el mercado laboral y la existencia de “techos de cristal”.

La participación económica evidencia una diferencia estructural importante. Solo el 38% de las mujeres forma parte de la PEA, frente a 62% en los hombres. La brecha de 24 puntos no se explica únicamente por decisiones individuales; también responde a restricciones estructurales como la falta de servicios de cuidado infantil, la distribución desigual del trabajo doméstico y el desempleo disfrazado. Además, el hecho de que haya más mujeres en la muestra total pero menos en la PEA indica una subutilización del potencial productivo femenino.

## 2.2 Comparación de la PEA

El análisis de la PEA forma parte de un subconjunto de nuestra muestra que reporta una vinculación activa con el mercado de trabajo, permite dilucidar no solo las condiciones generales de empleo en el país, sino también las asimetrías que atraviesa la participación por género.

Cuadro 2.2 PEA y estructura demográfica por sexo. ENOE 2019, 2021 y 2023.

	2019		2021		2023	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
*N	223,722	151,608	376,336	252,305	425,322	289,735
Edad Promedio	39.4	39.7	39.98	39.39	40.1	39.94
Salario	\$5162.06	\$3848.39	\$5764.16	\$4493	\$6649.8	\$5060.24
Casados	0.47	0.46	0.5	0.49	0.52	0.51
No sabe leer o escribir	0.03	0.02	0.02	0.02	0.02	0.02
Menor a Primaria	0.03	0.03	0.03	0.02	0.03	0.02
Primaria	0.21	0.17	0.2	0.17	0.18	0.15
Secundaria	0.31	0.31	0.3	0.27	0.29	0.27
Preparatoria	0.23	0.19	0.23	0.21	0.24	0.22
Universidad	0.22	0.31	0.24	0.33	0.26	0.33
Menor a 14 años	0.01	0	0.01	0.01	0.01	0.01
14 a 24 años	0.18	0.15	0.16	0.15	0.16	0.14
24 a 34 años	0.23	0.24	0.23	0.24	0.23	0.23
34 a 44 años	0.22	0.23	0.22	0.24	0.21	0.24
44 a 54 años	0.19	0.22	0.2	0.21	0.2	0.21
54 a 64 años	0.11	0.12	0.12	0.11	0.12	0.12
mayor a 65 años	0.06	0.04	0.06	0.04	0.06	0.04
Primarias	0.15	0.03	0.15	0.03	0.14	0.03
Secundarias	0.33	0.21	0.32	0.18	0.32	0.18
Terciarias	0.52	0.76	0.53	0.79	0.53	0.79
Seguridad social	0.41	0.41	0.4	0.42	0.4	0.4
**Nivel de ingreso						
Un salario mínimo	0.13	0.27	0.19	0.31	0.27	0.41
1-2 salarios mínimos	0.32	0.32	0.38	0.35	0.37	0.31
2-3 salarios mínimos	0.22	0.12	0.15	0.11	0.12	0.07
3-5 salarios mínimos	0.1	0.07	0.08	0.05	0.05	0.03
> 5 salarios mínimos	0.04	0.02	0.03	0.02	0.02	0.01
No especificado	0.19	0.19	0.18	0.17	0.18	0.18
Menos de 15 horas	0.04	0.13	0.04	0.11	0.04	0.11
15 a 34 horas	0.14	0.23	0.15	0.23	0.15	0.23
35 a 48 horas	0.47	0.43	0.46	0.44	0.47	0.45
Mas de 48 horas	0.32	0.18	0.32	0.18	0.31	0.18

Nota: \*Contempla únicamente a la población que indicó ser económicamente activa, y registró la duración de su jornada laboral. \*\* La diferencia del total con respecto a los registros de nivel de ingreso corresponden a las muestras que no reciben algún ingreso o se encuentra no es especificado.

Fuente: Creación propia con datos tomados de la ENOE (INEGI) 2019, 2021 y 2022

El cuadro 2 presenta la estadística descriptiva sobre las características de hombres y mujeres que entre 2019 y 2023, formaron parte de la fuerza laboral en México y registraron una jornada laboral definida. A partir de esta información, es posible observar más allá de los promedios evidentes, una serie de patrones estructurales que diferencian significativamente la experiencia laboral de mujeres y hombres, incluso cuando ambos sexos se encuentran formalmente integrados al mercado.

Una de las dimensiones más visibles de las diferencias entre los géneros en la PEA es el nivel salarial. A lo largo de los tres años analizados, el ingreso promedio mensual de los hombres supera consistentemente al de las mujeres. En 2019, los hombres obtuvieron una diferencia de más del 25% sobre las mujeres. Esta brecha no se reduce con el tiempo; por lo contrario, se amplía ligeramente para 2023, Esta diferencia no puede ser atribuida exclusivamente a características productivas observables, ya que, como se observará en los apartados siguientes, muchas mujeres en la PEA presentan niveles educativos comparables o incluso superiores. Por tanto, se sugiere que hay factores estructurales y de tipo discriminatorio.

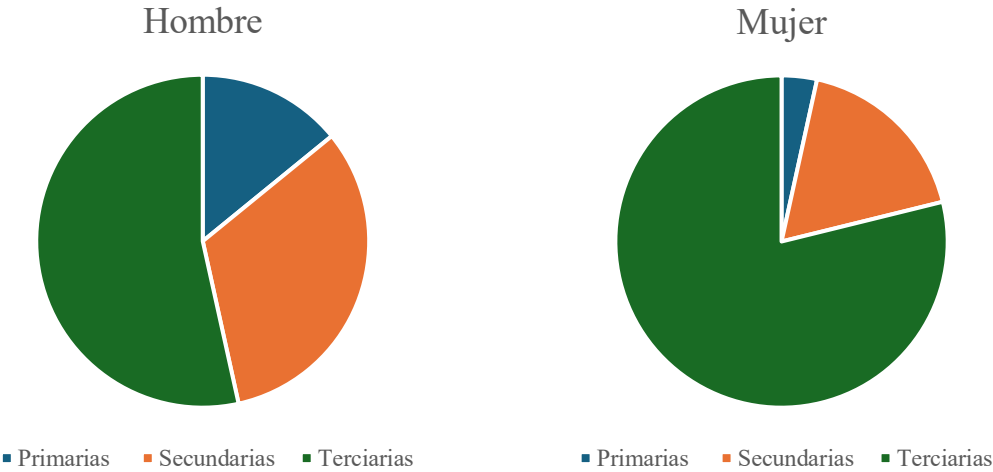
El promedio de edad para ambos géneros se mantiene relativamente estable y cercano entre sí: hombres con 40.1 años y mujeres con 39.94 en 2023. Sin embargo, este dato oculta trayectorias laborales divergentes. Que ambos promedios sean altos sugiere que el mercado laboral mexicano incorpora con mayor fuerza a personas en etapas medias de la vida y deja al margen tanto a jóvenes como a adultos mayores. Ahora bien, aunque los promedios por sexo se parezcan, el efecto de la edad sobre la experiencia no es igual. Las mujeres suelen iniciar su vida laboral más tarde y la interrumpen con mayor frecuencia por tareas de cuidado y maternidad. Esa dinámica distorsiona la acumulación de experiencia.

Una de las diferencias más claras aparece al desagregar por nivel educativo. Dentro de la población ocupada, las mujeres presentan una mayor proporción con estudios universitarios. En 2023, 33% de las mujeres ocupadas tenía educación superior, frente a 26% de los hombres. Esta distancia no se redujo en los años analizados, más bien se consolidó. Aún con mayor presencia en los tramos altos de escolaridad, las mujeres siguen recibiendo salarios más bajos. Esto sugiere ineficiencias en la asignación del capital humano y pérdidas de productividad a nivel agregado. Además, los hombres se concentran con mayor frecuencia en niveles intermedios, en particular

secundaria y media superior, mientras que las mujeres aparecen en los extremos, con más participación en universidad y un pequeño aumento en básica incompleta.

Existe una clara concentración sectorial de los trabajadores, esto es evidencia de una segregación horizontal, ya que cerca del 53% de los hombres se sitúa en actividades terciarias, mientras que las mujeres llegan a valores más altos, cerca del 80%. La participación masculina dentro de los primeros dos sectores es de 14 a 15% y 32 a 33% para secundarias, y para las mujeres su participación es mínima, esta segmentación está reflejada en los ingresos, ya que los sectores de mayor productividad suelen ser mejor pagados, con bonos, y remuneraciones mayores, ahí están situados los hombres. El peso de las mujeres en las actividades de cuidado, comercio, y servicios personales rara vez profesionalizados o bien remunerados, perpetúa esta brecha estructura.

Gráfico 2.1 Estructura sectorial de la PEA 2023



Fuente: Creación propia con datos de la ENOE 2023

El desglose por rango salarial revela una distribución claramente asimétrica. En 2023, el 41% de las mujeres percibían un salario mínimo en comparación de los hombres que presentaban una proporción del 27%, mientras asciende la escala de ingreso, la profundidad femenina se reduce abruptamente, donde cerca del 1% de las mujeres gana más de cinco salarios mínimos, en comparación del 2% de los hombres. En el rango de 3 a 5 salarios mínimos, la diferencia también es amplia (3% para mujeres vs 5% para hombres). Este patrón no ha cambiado en los tres años

analizados, lo que sugiere una estructura rígida con fuerte presencia del llamado “techo de cristal”, que impide el ascenso económico de las mujeres aún dentro de la PEA.

Al observar las horas trabajadas aparecen brechas evidentes. Para 2023, 23% de las mujeres ocupadas se ubicó entre 15 y 34 horas semanales, mientras que entre los hombres esa proporción fue 15%. También destaca que 11% de las mujeres trabajó menos de 15 horas, por encima del 4% registrado en los hombres. Esta inclinación hacia jornadas parciales puede leerse de dos maneras. Por un lado, como una estrategia individual de conciliación entre trabajo y familia. Por otro, y con mayor frecuencia, como una restricción estructural que empuja a aceptar empleos de medio tiempo, peor pagados y con menos derechos. Lejos de representar una flexibilidad moderna, este patrón configura un entorno de precariedad y de baja acumulación de derechos y de experiencia laboral.

En cuanto al acceso a seguridad social, se observa una paridad aparente: tanto hombres como mujeres presentan tasas estables entre 40% y 42% a lo largo del período. No obstante, esta aparente igualdad en cobertura oculta el hecho de que muchas mujeres activas se concentran en ocupaciones informales o con contratos inestables. La igualdad de proporciones en acceso a seguridad social no implica igualdad en la calidad del empleo: muchas mujeres podrían estar cubiertas por su pareja, por programas sociales o por instituciones no laborales. Además, la proporción de trabajadores sin especificación de nivel de ingreso se mantiene en torno al 18-19%. Lo que introduce un sesgo de opacidad que probablemente afecta más a mujeres con empleos informales o sin contrato.

La distribución por grupos de edad muestra que tanto hombres como mujeres están concentrados en tramos de 24 a 44 años, pero con una ligera mayor proporción de mujeres en los tramos de 34 a 44 años. Esto sugiere que la inserción femenina en el mercado laboral se da con mayor fuerza en etapas más avanzadas, quizá tras la reducción de la carga de cuidado de hijos pequeños. En los extremos de edad, sin embargo, la participación femenina se reduce: solo el 4% de las mujeres ocupadas tiene más de 65 años, frente a un 6% de los hombres. En el rango juvenil (14 a 24 años), las mujeres también están subrepresentadas, lo que evidencia un ciclo de entrada más tardío al mercado laboral, probablemente mediado por roles familiares y expectativas de género.

Estos datos identifican diferencias menos visibles, pero igualmente determinantes: trayectorias laborales más fragmentadas, condiciones más precarias, inserciones tardías, y estancamiento en la movilidad laboral para las mujeres. Aunque los promedios globales puedan parecer similares en algunos indicadores -como edad o acceso a seguridad social-, las estructuras internas revelan desigualdades profundas y persistentes. Uno de los hallazgos más consistentes es la disparidad en los ingresos: aun cuando las mujeres ocupadas presentan una mayor proporción con estudios universitarios, sus ingresos son sistemáticamente más bajos en todos los años analizados. A ello se suma su alta concentración en sectores terciarios y jornadas parciales, lo que sugiere una forma de integración laboral más precaria, posiblemente marcada por la necesidad de conciliar trabajo remunerado con labores domésticas.

La jornada laboral más reducida entre mujeres, la menor presencia en tramos altos de ingreso y su baja representación femenina en el mercado de trabajo se da en condiciones desventajosas, incluso cuando ya han superado la barrera de ingreso. Sin embargo, este cuadro, por su naturaleza, excluye aspectos fundamentales para la comprensión integral de la desigualdad. No refleja, por ejemplo:

- A las mujeres que desean trabajar, pero no lo hacen por barreras estructurales (falta de guarderías, transporte inseguro, roles tradicionales).
- El peso del trabajo no remunerado, especialmente el cuidado de menores, adultos mayores o personas con discapacidad.
- Factores cualitativos como discriminación en procesos de contratación, promoción, acoso laboral o segregación vertical.
- Brechas regionales o interseccionales (mujeres indígenas, rurales, o con discapacidad), que no se desagregan aquí, pero que pudieran profundizar las desigualdades observadas.

### 2.3 Primeras evidencias de brecha

La evidencia descriptiva muestra patrones consistentes de desigualdad salarial entre hombres y mujeres, especialmente en sectores con menor formalidad, menor escolaridad y regiones del sur del país. Estas observaciones justifican el uso de modelos econométricos que permitan estimar

los determinantes ajustados y no observables de la brecha salarial. A través de la estadística descriptiva de la muestra y algunos indicadores:

Cuadro 2.3 Brecha salarial bruta.

	2019		2021		2023	
	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
Salario	\$5,162.06	\$3,848.39	\$5,764.16	\$4,493.00	\$6,649.80	\$5,060.24
Δ%	25.45%		22.05%		23.90%	
Razón	74.55%		77.95%		76.10%	

Fuente: Creación propia, con datos de la ENOE, brecha mediante  $1 - (W_m/W_H)$ .

El comportamiento de la brecha salarial durante el período analizado refleja que prácticamente las mujeres percibían en promedio apenas tres cuartas partes del ingreso masculino, lo que se traduce en una brecha cercana al 25%, este comportamiento no solo revela la persistencia de desigualdades estructurales. Como mínimo efecto de la pandemia sobre la consistente brecha, pero también con aspectos de mayor profundidad, como la persistente segmentación ocupacional y las limitaciones de las políticas públicas para incidir de manera sostenida en la igualdad salarial. En este sentido, los resultados sugieren que cerrar la brecha salarial de género en México requiere no solo del crecimiento económico, sino de transformaciones institucionales y sociales de mayor alcance.

Cuadro 2.4 Relación por Rama económica

	2019	2021	2023	2019	2021	2023
	Hombre	Hombre	Hombre	Mujer	Mujer	Mujer
<i>Primarias</i>	14.8%	15.2%	14.1%	3.4%	3.3%	3.4%
<i>Secundarias</i>	32.8%	32.1%	32.4%	20.7%	17.8%	17.7%
<i>Terciarias</i>	52.5%	52.7%	53.5%	75.9%	78.9%	78.8%

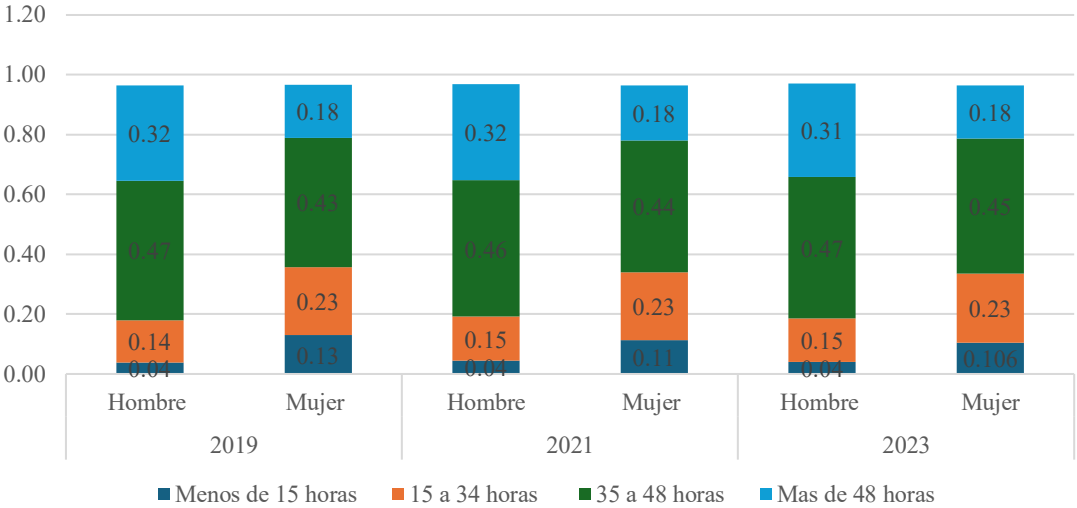
Fuente: Creación propia; tasa de participación de la PEA en ramas económicas

El cuadro muestra una segregación horizontal nítida: las mujeres se concentran en actividades servicios (terciarias), mientras que los hombres retienen mayor presencia en industria (secundarias) y primarias. Dado que las ramas secundarias suelen exhibir mayores niveles de productividad laboral y primas salariales que buena parte de los servicios personales y de comercio, esta composición arrastra diferencias en los ingresos medios por sexo aún antes de cualquier ajuste.

En términos dinámicos, la estructura sectorial es estable en el período, con leves oscilaciones al ciclo y a la reapertura post-pandemia. El hecho estilizado que interesa aquí es que la mayor densidad femenina en servicios coexiste con menor presencia en ramas con salarios promedios más altos, lo cual ofrece una primera pista de porque la brecha persiste incluso cuando los promedios educativos no son desfavorables a las mujeres.

El gráfico 2.2 ilustra visualmente esa asimetría de composición; el bloque de servicios domina el total femenino, mientras que en los hombres se observa una triada más equilibrada (primario-secundario-terciario). Este gráfico es clave para diferencias ingresos “en la puerta de entrada”, por el tipo de ocupaciones y ramas en las que efectivamente se insertan mujeres y hombres.

Grafico 2.2 Relación de la jornada laboral para PEA 2019, 2021 y 2023

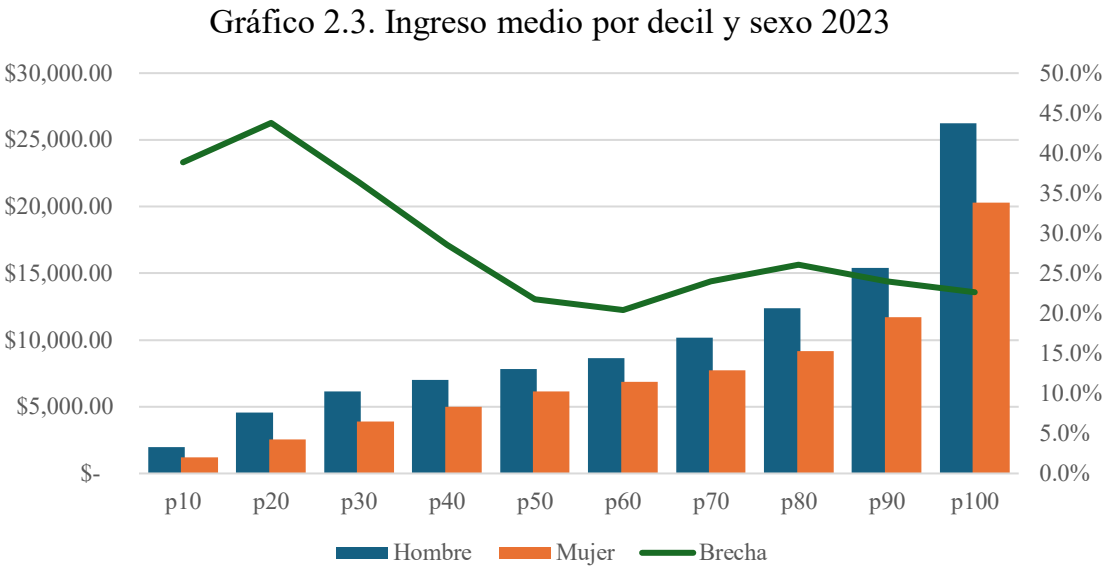


Fuente: Creación propia con datos de la ENOE, filtrado por PEA.

El gráfico resume de manera inmediata el sesgo hacia la parcialidad en el empleo femenino. (i) la modalidad modal en ambos sexos es 35-48 horas, pero las colas difieren; (ii) la densidad femenina en parciales es mayor, lo que sugiere que contribuye mecánicamente a menores ingresos aún a igual salario por hora.

Los datos de participación y tiempo trabajado confirman asimetrías en el margen extensivo e intensivo. En el extensivo, la PEA masculina supera de forma consistente a la femenina, lo que indica subutilización del potencial laboral de las mujeres. Es el intensivo, la distribución de horas es desigual: las mujeres registran mayor probabilidad de ubicarse en jornadas parciales (<15 y 15-34 horas), mientras que los hombres predominan en jornadas largas ( $\geq 49$  hrs).

Estas diferencias no solo afectan el ingreso mensual por masa de horas, sino que suelen ir de la mano de menor cobertura de derechos y menores trayectorias de acumulación (antigüedad, bonos, escalafone). Este grafico sugiere que la brecha salarial bruta no proviene únicamente de “precios” diferentes, sino también de intensidades de trabajo y disponibilidad horaria.



Fuente: Creación propia con datos de la ENOE, únicamente PEA, con método de winsorización p1, p99, y brecha salarial con  $1 - (W_M/W_H)$

El grafico 2.3 muestra una brecha en todos los deciles, los hombres en promedio perciben mayores ingresos que las mujeres. Esto confirma que la desigualdad salarial no es un fenómeno aislado de grupos en particulares. En los primeros deciles, los ingresos de mujeres y hombres son reducidos, sin embargo, la brecha relativa es particularmente elevada, superando algunos casos el 40%, este hallazgo sugiere que las mujeres enfrentan mayores dificultades para escapar de los segmentos de menor remuneración, incluso cuando su participación laboral es similar o

superior. Este fenómeno es conocido en la literatura como “piso pegajoso”, el cual las trabajadoras permanecen atrapadas en ocupaciones de baja productividad, con poca movilidad, ascendente y limitadas protecciones sociales.

En el extremo superior de la distribución (decil p90 y p100), aunque los ingresos son más elevados para ambos géneros, se mantiene una diferencia significativa. La presencia femenina en los niveles jerárquicos altos es proporcionalmente menor y, aunque sus ingresos superan el promedio general, se mantienen por debajo de los de sus pares hombres. Este fenómeno, conocido como techo de cristal, alude a un conjunto de barreras invisibles que limita el acceso de las mujeres a puestos de mayor nivel y mejor remuneración, aun cuando cuentan con credenciales educativas y experiencia comparables. Este gráfico del 2023 tiene un patrón muy similar a la distribución por deciles de los años 2021 y 2019, donde los extremos de la distribución son notablemente más pronunciados.

Estos hallazgos tienen importantes implicaciones. La existencia de brechas tanto en la base como en la cúspide salarial sugiere que las políticas públicas deben atender no solo la inserción laboral femenina en empleos de calidad, sino también la movilidad dentro del mercado de trabajo.

Estos estadísticos son consistentes y dejan en claro la existencia de brecha salarial bruta persistente a favor de los hombres a lo largo del período, así como la composición sectorial es asimétrica, ya que ellas están sobre representadas en servicios, ellos en actividades primarias-secundarias con primas salariales históricamente mayores, y por tanto menor masa salarial mensual; y la heterogeneidad territorial de la brecha es amplia. Todo ello sugiere que una parte importante de la desigualdad se genera incluso antes de considerar retornos condicionados; por eso los siguientes capítulos estimaran descomposiciones y correcciones de selección para separar composiciones de retornos. Como referencia de contexto, los organismos internacionales han documentado brechas ajustadas relevantes y con alta variación entre países, a menudo medidas sobre la mediana de ingresos de tiempo completo, lo que apunta a problemas de medición y comparabilidad que este trabajo toma en cuenta al fijar su estrategia empírica.

Este capítulo tiene como finalidad ofrecer un primer acercamiento a la realidad laboral de ambos géneros, en México, empleando como insumo central la Encuesta Nacional de Ocupación y Empleo (ENOE). A través de esta fuente, se logró condensar un conjunto de estadísticas básicas lo que la realidad es una encuesta de enorme amplitud, con cientos de variables y millones de observaciones que representan de manera fiel a la población mexicana.

De entre toda esa riqueza informativa, se eligieron únicamente las variables más vinculadas con el tema de investigación: educación, edad, estado civil, formalidad, rama de actividad, ingreso y sexo. La decisión de acotar las variables no busca excluir por excluir. Responde a la necesidad de trabajar con los factores que la teoría económica y la evidencia empírica identifican como determinantes de la brecha salarial. De este modo se evita dispersarse en indicadores secundarios y se concentra la atención en lo realmente relevante para explicar las diferencias de ingreso entre mujeres y hombres.

Es importante señalar que, pese a la depuración realizada, la base mantiene la representatividad nacional de la ENOE. Dicho de manera simple, la muestra refleja de forma fiel a la población de 15 años y más que participa en el mercado de trabajo, lo que da confianza en que las tendencias detectadas no son un artefacto de la selección de casos. Así, este capítulo no solo describe el panorama, también legitima el análisis empírico posterior.

Los resultados apuntan a una constante: las mujeres reciben ingresos menores en prácticamente todos los deciles. De ahí surgen dos figuras muy discutidas en la literatura, el piso pegajoso y el techo de cristal. La primera retiene a las trabajadoras en ocupaciones mal pagadas y con escasos derechos, la segunda limita su llegada a puestos de alta responsabilidad y mejor salario. Con estas piezas sobre la mesa, todo indica que la desigualdad responde a un mecanismo persistente y no a fluctuaciones aleatorias dentro de la distribución de salarios.

En síntesis, este capítulo cumple un doble propósito. Primero, ofrece un panorama general de la muestra y de sus rasgos más relevantes. Segundo, muestra indicios claros de que la brecha salarial de género persiste aun cuando han mejorado las condiciones educativas y la participación laboral de las mujeres. La comparación entre 2019, 2021 y 2023 confirma que el fenómeno se mantuvo antes, durante y después de la pandemia, con variaciones en su magnitud.

Este recorrido descriptivo, breve pero sustantivo, abre la puerta al análisis econométrico que se desarrollará en los capítulos siguientes. Si hasta aquí hemos mostrado el problema en cifras y gráficos, el paso que sigue será explicar sus causas con mayor rigor, descomponiendo los efectos atribuibles a características observables y aquellos que responden a mecanismos más profundos de discriminación.

## CAPÍTULO III. Metodología<sup>2</sup>

Se ha utilizado la descomposición Oaxaca-Blinder en su forma doble y triple, para el análisis de la diferencia salarial entre hombres y mujeres, también se implementó la corrección en selección de sesgo Heckman, para estimar la función separada de cada grupo de observaciones donde también se generó un modelo agrupado (pooled) como referencia no discriminatoria, y para calcular la robustez de resultados se emplearon técnicas de remuestreo Bootstrap.

El objetivo central de la descomposición O-B es identificar qué proporción de la diferencia en los resultados salariales entre hombres y mujeres se debe a la diferencia en los niveles de las variables explicativas y que parte se debe a la diferencia en los niveles de las variables explicativas.

### 3.1 Observaciones técnicas

#### 3.1.1 Construcción previa a los modelos, exclusión y conformación de las variables

Definición de la muestra y filtros básicos.

Este estudio utiliza microdatos de la ENOE 2019, 2021 y 2023 construye y pool de cortes transversales. El dominio analítico es la PEA ocupada con ingreso laboral observado. A continuación, se documentan, con precisión econométrica, los filtros y transformaciones iniciales.

##### a) Entrevista completa

Se observa únicamente a las personas con cuestionarios concluidos. Definimos el indicador de completitud:

$$\kappa_i = \mathbb{1}(r\_def_i = 0)$$

y restringimos la base a  $\kappa_i = 1$ . En notación de conjunto:

---

<sup>2</sup> Reproducibilidad. Todos los do-files y scripts utilizados para la preparación de datos, estimaciones y gráficos se incluyen en los **Anexos B1, C, D**, junto con los metadatos y versiones de software (STATA 17; R 2022.12.0+353 con paquetes *oaxaca*, *quantreg*, *sampleSelection*, entre otros). Las definiciones de variables, ponderadores y criterios muestrales constan en el **Anexo A y B**.

$$\mathcal{S}_0 = \{i: \kappa_i = 1\}$$

b) Edad y grupos etarios

Se estandariza la variable de edad y se crean participaciones exhaustivas y mutuamente excluyentes:

$$edad_i \in \mathbb{N}, \quad g_i = \begin{cases} 1 & \text{si } edad_i \leq 14, \\ 2 & \text{si } 15 \leq edad_i \leq 24 \\ 3 & \text{si } 25 \leq edad_i \leq 34 \\ 4 & \text{si } 35 \leq edad_i \leq 44 \\ 5 & \text{si } 45 \leq edad_i \leq 54 \\ 6 & \text{si } 55 \leq edad_i \leq 64 \\ 7 & \text{si } edad_i \geq 65 \end{cases}$$

Para los ejercicios econométricos se utilizarán dummies  $D_{ig} = \mathbb{1}(g_i = g)$ ,  $g \in \{1, \dots, 7\}$  con la categoría base que se indique en cada especificación.

c) Con la codificación oficial de ENOE se define:

$$ocupado_i = \mathbb{1}(clase3_i = 1),$$

Que identifica ocupación remunerada. La muestra se recorta a:

$$\mathcal{S}_1 = \{i \in \mathcal{S}_0: ocupado_i = 1\}$$

Para asegurar observaciones laborales válidas, se aplican recortes de consistencia (p. ej., valores admisibles en  $clase1_i$  y rama de actividad  $rama\_est2_i$ ), eliminando registros con claves faltantes o no clasificables.

$$\mathcal{S}_2 = \{i \in \mathcal{S}_1: clase1_i \in \mathcal{A}, rama\_est2_i \in \mathcal{R}\}$$

Donde  $\mathcal{A}$  y  $\mathcal{R}$  son los subconjuntos admisibles según el diccionario de ENOE

d) Ingreso laboral y transformación logarítmica

El salario monetario se toma como el ingreso por trabajo principal observado ( $ingocup_i$ ). La variable dependiente se define como:

$$y_i = \ln(\text{ingocup}_i),$$

Con la restricción  $\text{ingocup}_i > 0$ . Esta transformación reduce asimetría, valores extremos y mejora las propiedades estadísticas. El universo efectivo para el análisis de salario queda:

$$\mathcal{S} = \{i \in \mathcal{S}_2: \text{ingocup}_i > 0\}^3$$

Posteriormente construyeron variables explicativas, el vector de controles  $x_i$  para capturar capital humano, estructura familiar, sector de actividad y rasgos territoriales:

Capital humano y habilidades básicas

- Sabe\_leer\_escribir  $\leftarrow$  cs\_p12.
- grado\_esc (máximo grado alcanzado) y años\_esc (cs\_p13\_1 y cs\_p13\_2), además de una versión categórica grado\_escolar (Menor a primaria, Primaria, Secundaria, Preparatoria, Profesional).

Estructura familiar

- n\_hijos3 = 1 si reporta 1–3 hijos; n\_hijos6 = 1 si reporta 4–25 hijos.
- Estado civil: e\_civil = 1 casado (e\_con==1 o 6); 0 en otros estados.

Protección social y servicios

- seg\_soc (acceso a seguridad social) y serv\_med a partir de medica5c (tiene/no tiene acceso).

Estructura sectorial y rama

- rama\_est1 para clasificar: genera dummies ActPrim, ActSec, ActTer.

---

<sup>3</sup> Hasta este punto, la base queda condicionada a PEA ocupada con ingreso reportado, que es el subconjunto relevante para estimar funciones de ingresos y descomposiciones salariales. Salvo especificación contraria, las estadísticas descriptivas y estimaciones utilizarán los factores de expansión provistos por ENOE ( $w_i$ ), de modo que cualquier promedio  $\bar{z}$  se calcula como:  $\bar{z} = \frac{\sum_{i \in \mathcal{S}} w_i z_i}{\sum_{i \in \mathcal{S}} w_i}$ .

Duración/tiempo y otras

- Variables de duración `dur_est`, categorías de ingreso `ing7c`, zona `zonaA/zonaB`, `ur` (urbano/rural).

### 3.1.2 Ponderador

La ENOE es una encuesta probabilística con factores de expansión que permiten inflar cada observación a su población objetivo. En el do-file se declara un diseño “solo pesos” con `svyset [pw=est]` y, en estimaciones OLS por subgrupos, se emplea también `[pw=fac]`. A continuación, se formalizan los estimadores y sus varianzas bajo ponderación.

#### a) Factor de expansión y normalización

Sea  $w_i > 0$  el peso de ENOE (`est/fac`). Para el análisis por año  $t$  se usan los pesos originales; para el pool 2019-2023 se puede re-escalar para que cada año contribuya con población comparable:

$$w_i^* = w_i \cdot \frac{\bar{N}}{N_t}, \quad N_t = \sum_{i \in S_t} w_i, \quad \bar{N} = \frac{1}{T} \sum_{t=1}^T N_t$$

$w_i$  es la población expandida del año y  $\bar{N}$  su media interanual, a fin de evitar que un año domine por tamaño muestral. Se declaro un diseño ‘solo pesos’ con `svyset [pweight=est]`, por lo que las medias, proporciones y regresiones se estimaron con `p-weights` y errores estándar por linealización. En algunas regresiones descriptivas por subgrupo (sexo-Año) se usó OLS ponderado con el peso original del año (`[pweight=fac]`), reportando errores estándar robustos.

Con pesos, las medias poblacionales se computan como:  $\bar{z} = \frac{\sum_i w_i z_i}{\sum_i w_i}$ , y los modelos lineales se estiman minimizando  $\sum_i w_i (y_i - x_i' \beta)^2$ , garantizando indicadores a nivel poblacional.

Nota previa

En todos los modelos estimados se incorporó el diseño muestral complejo de la ENOE mediante ponderadores de expansión y estratificación. Para las ecuaciones salariales tipo Mincer y las

descomposiciones Oaxaca-Blinder se emplearon ponderadores de probabilidad y errores estándar robustos a heterocedasticidad. En el caso de la descomposición (doble y triple) se obtuvieron errores estándar y bandas de confianza mediante re-muestreo bootstrap, lo que permite evaluar la significancia estadística de los componentes  $E$ ,  $C$  e  $I$ . Para el modelo de corrección de selección de Heckman se utilizó el estimador de dos etapas con errores estándar robustos, y para el modelo espacial SLX en panel se estimaron regresiones con efectos fijos municipales y temporales, reportando errores estándar agrupados a nivel municipal, de manera consistente con la literatura de paneles espaciales.

### 3.2 Ecuación salarial (Mincer) por sexo

Estimar estas funciones de salarios separados por sexo para obtener retornos a la escolaridad y premios o penalizaciones asociados a características demográficas, familiares y sectoriales. Estas estimaciones ayudarán en la descomposición Oaxaca Blinder.

Previamente ya se ha modelado el logaritmo del ingreso laboral observado del empleo principal, mismo que permite interpretar coeficientes como semielasticidades y atenuar asimetrías de la distribución. Se ejecuta la estimación por sexo usando “by sex:” para obtener conjuntos de parámetros  $\beta_H$  y  $\beta_M$  sobre las mismas covariables. Con ello se evita imponer que hombres y mujeres compartan la misma función salarial.

De igual manera se usa el peso original de ENOE a nivel persona ([pw=fac]) de esta forma, los coeficientes representan promedios poblacionales y los grupos con mayor representación demográfica (según ENOE) contribuyen en proporción a su tamaño.

La ecuación salarial de referencia sigue la forma ampliada de Mincer, estimada por separada por hombres y mujeres. Sea  $y_i$  el logaritmo natural del ingreso laboral mensual observado del individuo  $i$ , y  $X_i$  el vector de características socioeconómicas (escolaridad, grupos de edad, estado civil, condición de formalidad, rama de actividad, entre otras). Para cada sexo  $g \in \{hombres, mujeres\}$ , se especifica:

$$y_{ig} = \beta_{0g} + X'_{ig}\beta_g + \varepsilon_{ig}$$

Donde  $\beta_{0g}$  es la ordenada al origen,  $\beta_g$  el vector de coeficientes asociados a las covariables y  $\varepsilon_{ig}$  el término de error, con  $\mathbb{E}(\varepsilon_{ig}|X_{ig}) = 0$ . En notación matricial, para cada grupo  $g$  se tiene:

$$y_g = X_g\beta_g + \varepsilon_g$$

Donde  $y_g$  es el vector  $n_g * 1$  de log-ingresos,  $X_g$  es la matriz  $n_g * k$  de covariables y  $\varepsilon_g$  el vector de perturbación aleatorias. Esta especificación es la base para la posterior descomposición Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973).

La estimación de la ecuación Mincer ampliada para cada sexo se estima:

$$y_i = \alpha_s + \beta_s ur_i + \gamma'_s G_i + \lambda_s Sabe\_leer\_escribir_i + \theta'_s E_i + \rho_s años\_esc_i + \mu_s e\_civil_i \\ + \sigma_s seg\_soc_i + \zeta'_s R_i + \kappa'_s C_i + \tau_s dur\_est_i + \chi_s serv\_med_i + \nu'_s N_i + \varepsilon_{is}$$

Donde:

- *ur*: Identifica zona urbana
- $G_i$ : Dummies de grupo de edad (una categoría base)
- *Sabe\_leer\_escribir*: Indicador de alfabetización
- $E_i$ : Dummies de grado educativo alcanzado (*grado\_esc*, una base)
- *años\_esc*: años completos de escolaridad (se incluye para capturar el retorno marginal dentro de niveles).
- *e\_civil*: Casado/a.
- *seg\_soc*: acceso a seguridad social.
- $R_i$ : Dummies de rama/sector económico (*rama*, una base).
- $C_i$ : Dummies de categoría de ingreso (*ing7c*, una base).
- *dur\_est*: Clasificación de la población ocupada por duración de la jornada laboral
- $N_i$ : Dummies de nivel de instrucción (*niv\_ins*, una base)<sup>4</sup>.

Se usó MCO ponderado con p-weights ( $[pw = fac]$ ). En notación matricial:

---

<sup>4</sup> En todos los bloques de dummies se omitió una referencia: rural ( $ur=0$ ); 25–34 años (*grupo\_edad*); Secundaria (*grado\_esc*); Terciario/Servicios (*rama*); c1 en *ing7c* (si se incluye); y Básico en *niv\_ins*. En dummies binarias, la base es el valor 0 (no alfabetizado, soltero/a, sin seguridad social, sin servicio médico). Las variables continuas (*años\_esc*, *dur\_est*) no requieren base.

$$\hat{\beta}_s = \arg \min_{\beta_s} \sum_{i \in S_s} w_i (y_i - x_i' \beta_s)^2, \quad w_i = fac_i$$

De tal manera que la regresión Mincer ampliada del log-salario se aplica por cada sexo, sobre urbanidad, grupos de edad, alfabetización, nivel educativo, y años de escolaridad, estado civil, seguridad social, rama de actividad categorías de ingreso, duración en el empleo, acceso a servicios médicos y nivel de instrucción; todo ello ponderado con el factor de expansión de la ENOE.

### 3.2 Mincer Base (por sexo) y Oaxaca Blinder

Descomposición Oaxaca-Blinder con diseño de encuesta  $g \in \{m, w\}$  los grupos de sexo (hombres, mujeres). La variable dependiente es:

$$y_i = \ln(\text{ingocup}_i) = \text{lsalario}_i$$

El vector de covariables para cada persona  $i$  es:

$$\mathbf{X}'_i = [1, ur_i, \mathbf{G}_i, Sabe_i, \mathbf{E}_i, \text{años\_esc}_i, e\_civil_i, seg\_soc_i, \mathbf{R}_i, \mathbf{C}_i, dur\_est_i, serv\_med_i, \mathbf{N}_i]$$

Donde  $(\mathbf{G}_i)$  (grupos de edad),  $(\mathbf{E}_i)$  (grado\_esc),  $(\mathbf{R}_i)$  (rama),  $(\mathbf{C}_i)$  (ing7c) y  $(\mathbf{N}_i)$  (niv\_ins) son dummies con categoría base omitida.

Con los pesos de ENOE previamente declarados en svyset<sup>5</sup> (p-weights) denotamos  $w_i > 0$ , las funciones salariales por sexo se estiman por MCO ponderado:

$$\hat{\beta}_g = \arg \min_{\beta_g} \sum_{i \in S_g} w_i (y_i - X_i' \beta_g)^2, \quad g \in \{m, w\}$$

Los promedios ponderados (poblacionales) son:

$$\bar{y}_g = \frac{\sum_{i \in S_g} w_i y_i}{\sum_{i \in S_g} w_i}, \quad \bar{X}_g = \frac{\sum_{i \in S_g} w_i X_i}{\sum_{i \in S_g} w_i}$$

<sup>5</sup> Con la opción svy, Oaxaca usa el diseño de encuesta declarado para obtener errores estándar por linealización de  $E$ ,  $U$  y sus desgloses. Los coeficientes y medias ponderadas se calculan con  $w_i$

La brecha salarial que observamos entre hombres y mujeres no solo una cosa: es la suma de (i) como están distribuidas sus características (edad), educación formalidad sector, etc.) y (ii) como se “pagan” esas mismas características en el mercado para cada sexo. La descomposición Oaxaca Blinder toma esta idea y traduce en una contabilidad clara: Una parte explicada por composición y otra por retornos.

$$\Delta = \bar{y}_m - \bar{y}_w = \bar{X}'_m \hat{\beta}_m - \bar{X}'_w \hat{\beta}_w \quad 6$$

Aquí  $\bar{X}_g$  son los promedios ponderados de las características para cada sexo  $g \in \{m, w\}$  y  $\hat{\beta}_g$  son los coeficientes (retornos) estimados en la Mincer por sexo. Esta identidad dice: la brecha total es lo que obtienes si evalúas “las características promedio de cada grupo” con los precios de su propio grupo.

Ahora bien, esa brecha  $\Delta$  aún una mezcla de composición y precios. Para separar ambas cosas introducimos una estructura de referencia  $\beta^R$ , es un árbitro que fija un punto de precios común: puede ser el de hombres o el de mujeres o uno no discriminatorio, estimado en la muestra combinada (pooled) <sup>7</sup>. Con ese libro de precios en la mano, reescribimos  $\Delta$  sumando y restando los mismos términos

$$\Delta = \underbrace{(\bar{X}_m - \bar{X}_w)' \beta^R}_{\text{Composición (E)}} + \underbrace{\bar{X}'_m (\hat{\beta}_m - \beta^R) + \bar{X}'_w (\beta^R - \hat{\beta}_w)}_{\text{Retornos (U)}}$$

- Composición (E): ¿Qué parte de la brecha se explica porque los hombres y las mujeres no tienen las mismas características promedio? Para contestarlo, mantenemos fijo el “libro de precios”  $\beta^R$  y simplemente comparamos medias:  $(\bar{X}_m - \bar{X}_w)' \beta^R$ . Si los promedios fueran idénticos, este término sería cero.
- Retornos(U): Aún si igualaras las medias, puede quedar brecha por que los precios no son los mismos para cada sexo. Este residuo se captura con:

---

<sup>6</sup> Con svy, los errores estándar de todos los componentes se obtienen por linealización de Taylor bajo el diseño de “solo pesos”.

<sup>7</sup> El estimador Pooled consiste en estimar una ecuación salarial única para toda la muestra (hombres y mujeres combinados), cuyos coeficientes se interpretan como una “estructura salarial no discriminatoria” de referencia. Bajo este esquema, las dotaciones de cada grupo se valoran con retornos promedio de la muestra conjunta, evitando el sesgo de tomar exclusivamente la estructura de un grupo como referencia, por ejemplo, masculina.

$$\bar{X}'_m(\hat{\beta}_m - \beta^R) + \bar{X}'_w(\beta^R - \hat{\beta}_w)$$

Que mide cuanto se alejan los precios propios de cada grupo de los precios de referencia, ponderados por sus medias. De manera contrafactual la podemos observar como “si ambos jugaran en el mismo libro de precios”  $\beta^R$ .

$$\bar{y}_m^R = \bar{X}'_m \beta^R, \quad \bar{y}_w^R = \bar{X}'_w \beta^R$$

Entonces,  $E = \bar{y}_m^R - \bar{y}_w^R$  responde, “¿Cuánta brecha quedaría si los precios fueran los de referencia y lo único distinto fueran las dotaciones?” y  $U$  suma el ajuste de precios de cada grupo: cuanto cambian  $\bar{y}_m^R$  y  $\bar{y}_w^R$  cuando los evaluamos con sus propios precios  $\hat{\beta}_m$  y  $\hat{\beta}_w$ .<sup>8</sup>

La manera, por lo tanto, de interpretar los valores porcentuales tomando en cuenta el salario a partir de su expresión *log – salarios*, la traducción a porcentajes es directa y más intuitiva:

$$R = \exp(\Delta), \quad R_E = \exp(E), \quad R_U = \exp(U), \quad R \approx R_E \cdot R_U.$$

- $R - 1$  (en %) es la brecha total en niveles
- $R_E - 1$  (en %) es la parte explicada por composición
- $R_U - 1$  (en %) es la parte no explicada (retornos)<sup>9 10</sup>

En suma:  $E$  responde a “quién es quién” (dotaciones) y  $U$  a “Cuanto paga el mercado” (precios). La metodología separa ambas caras de la brecha, une ponderadores de encuesta para respetar la población objetivo y permite atribuir con precisión la brecha total a bloques y variables concretas. Con eso, los resultados dejan de ser una sola cifra única y se convierten en un mapa de donde está la brecha y por qué.

---

<sup>8</sup> Contribuciones por variable  $j$ . Así, cuanto de  $E$  proviene de educación, cuanto, de edad, etc.; y cuanto de  $U$  se asocia a diferencias de precios en cada covariable:

$$E_j = (\bar{X}_{mj} - \bar{X}_{wj})\beta_j^R, \quad U_{mj} = \bar{X}_{mj}(\hat{\beta}_{mj} - \beta_j^R), \quad U_{wj} = \bar{X}_{wj}(\hat{\beta}_{wj} - \beta_j^R),$$

$$E = \sum_j E_j, \quad U = \sum_j (U_{mj} + U_{wj}).$$

Esto es clave para políticas públicas si  $E_{\text{educación}}$  es grande y positivo, la brecha se explica por una diferente distribución educativa; si  $U_{\text{educación}}$  es grande, a igual educación, el retorno se la educación difiere entre sexos

<sup>9</sup> Si lo promedios de características fueran iguales,  $E = 0$ : todo sería retornos.

<sup>10</sup> Si los precios fueran iguales ( $\hat{\beta}_m = \hat{\beta}_w = \beta^R$ ),  $U = 0$ : todo sería composición.

### 3.3 Variante con corrección de selección (Heckman)

En nuestra base solo vemos salarios de quienes están ocupados, Si la decisión de participar (trabajar) no es aleatoria – Por ejemplo, las mujeres con niños pequeños o con menor capital social podrían salir al mercado pese a tener productividad alta o baja. En este caso, la regresión mincerina por sexo hereda un sesgo de selección porque el término de errores del salario está correlacionado con el término que gobierna la participación. Heckman corrige justamente eso al modelar dos ecuaciones: una selección (participación) y otra de salarios, vinculadas por una correlación  $\rho$ . La correlación aparece como una variable adicional, la razón de Mills inversa  $\lambda$ , que “desvía” el sesgo.

Dado que el salario solo se observa para quienes están ocupados y reportan ingreso, se corrige el posible sesgo de selección mediante el modelo de dos etapas propuesto por Heckman (1979). El procedimiento parte de dos ecuaciones: una de selección al empleo remunerado y otra de resultados salariales.

En la primera etapa, se modela la probabilidad de observar salario mediante una ecuación de selección tipo probit:

$$d_i^* = Z_i' \gamma + u_i, \quad d_i \begin{cases} 1 & \text{si } d_i^* > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Donde  $d_i$  indica participación con salario observado,  $Z_i$  es el vector de variables que afectan la participación (incluye un subconjunto de  $X_i$  y cuando es posible, instrumentos de exclusión) y  $u_i$  es el término de error.

Modelo formal, dos ecuaciones

Selección (participación laboral):

$$s_i^* = z_i' \gamma + u_i, \quad s_i = 1(s_i^* > 0)$$

Resultados (salario observado solo si  $s_i = 1$ ):

$$y_i = \ln(\text{ingocup}_i) = x_i' \beta + \varepsilon_i$$

Supuesto clave de Heckman (normal bivariada)

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{bmatrix} \right)$$

Bajo este supuesto, el salario esperado condicionado a estar ocupado es:

$$E[y_i | x_i, z_i, s_i = 1] = x_i'\beta + \delta\lambda(z_i'\gamma), \quad \delta \equiv \rho\sigma_\varepsilon, \quad \lambda(t) \equiv \frac{\phi(t)}{\Phi(t)}.$$

Si  $\delta = 0$  (equivalente  $\rho = 0$ ), no hay sesgo de selección.

Para identificar que el modelo este bien identificado en la ecuación de selecciones incluimos variables que afectan la probabilidad de ocupar un ejemplo, pero que, condicionado el resto de los controles, no entran en la ecuación de salarios. En el modelo de selección se incluyen variables como Sabe\_leer\_escribir, niv\_ins, n\_hijos3 y n\_hijos6, ya que permiten capturar aspectos de alfabetismo, nivel educativo básico y factores familiares que condicionan la empleabilidad y la decisión de participar en el mercado laboral. En particular, el alfabetismo y el nivel de instrucción inicial inciden en las probabilidades de inserción y búsqueda de empleo, mientras que el “precio” del capital humano se refleja directamente en los salarios a través de variables como grado\_esc (dummies) y años\_esc (continuo). Asimismo, el número de hijos tiene un efecto relevante sobre la participación laboral al representar costos de oportunidad y tiempos de cuidado, pero no constituye un determinante directo del salario una vez que se controlan otros factores como edad, educación, formalidad y rama de actividad económica.

Conformada la estimación con pesos ENOE

$$\hat{\gamma} = \arg \max_{\gamma} \sum_i [s_i \ln \Phi(z_i'\gamma) + (1 - s_i) \ln(1 - \Phi(z_i'\gamma))],$$

Usando p-weights  $w_i$

$$t_i = z_i'\hat{\gamma}, \quad \hat{\lambda}_i = \frac{\phi(t_i)}{\Phi(t_i)} \quad \text{si } s_i = 1. \text{ }^{11}$$

Posteriormente añadimos  $\hat{\lambda}_i$  sobre ocupados y por sexo  $s \in \{H, M\}$  como regresor:

---

<sup>11</sup>  $\hat{\lambda}_i$ : si  $\Phi(t_i)$  (probabilidad de estar ocupado) es baja,  $\hat{\lambda}_i$  es alta. Es decir, observaciones “poco probables” de estar ocupados reciben más corrección.

$$y_i = x_i' \beta_s + \delta_s \hat{\lambda}_i + \varepsilon_{is}, \quad i: s_i = 1.$$

- $\delta_s$  es la intensidad del sesgo en el grupo  $s$

Estimador  $\rho$  para la correlación entre errores de la ecuación de selección (probabilidad de estar ocupado) y de la ecuación de salarios.

- $\hat{\rho} > 0 \Rightarrow$  Selección positiva: Los tienen, en promedio, mayor salario potencial que los no ocupados (los no ocupados están “por debajo” en el componente no observado).
- $\hat{\rho} < 0 \Rightarrow$  Selección negativa: entre los no ocupados hay individuos con alto salarios potenciales que no participan (barreras, cuidados, etc.).<sup>12</sup>

Corrección en el análisis

Restamos la parte de selección estimada y reestimamos todo con ese salario:

$$y_i^{corr} = y_i - \hat{\delta}_s \hat{\lambda}_i$$

Con  $y^{corr}$  repetimos la Mincer por sexo (mismos  $X$  y pesos) y alimentamos la estimación de Oaxaca-Blinder:

$$\Delta^{corr} = y_m^{-corr} - y_w^{-corr} = E^{corr} + U^{corr}$$

Mantenemos  $\hat{\lambda}_i$  como regresor usamos  $y_i$  original y dejamos  $\hat{\lambda}_i$  dentro de la Estimación Mincer y de la DO-B. Esto nos ayuda a tener un bloque de tablas adicionales estimadas.

Descomposición Oaxaca Blinder doble y triple (explicación previa)

Sea  $\bar{y}_h$  el logaritmo promedio del salario para hombres y  $\bar{y}_m$  el correspondiente para mujeres. La brecha salarial total en logaritmos se define como:

$$\Delta = \bar{y}_h - \bar{y}_m$$

Denotando por  $\bar{X}_h$  y  $\bar{X}_m$  los vectores de medias de covariables, y por  $\hat{\beta}_h$  y  $\hat{\beta}_m$  los coeficientes estimados en cada grupo, la descomposición Oaxaca-Blinder clásica (doble) utiliza un vector de referencia  $\hat{\beta}^*$  y escribe:

---

<sup>12</sup> En la práctica Stata reporta  $athrho$ ;  $\hat{\rho} = \tanh(\widehat{athrho})$

$$\Delta = (\bar{X}_h - \bar{X}_m)' + \bar{X}_h'(\hat{\beta}_h - \hat{\beta}^*) + \bar{X}_m'(\hat{\beta}^* - \hat{\beta}_h)$$

Donde el primer término corresponde al componente explicado por dotaciones y los términos restantes al componente no explicado por retornos (Oaxaca, 1973; Blinder, 1973).

Siguiendo la propuesta de Jann (2008), la parte no explicada se puede desagregar en un componente de cambios en retornos puros (C) y un componente de interacción (I), dando lugar a la descomposición triple:

$$\Delta = E + C + I$$

Donde *E* recoge diferencias en características, *C* diferencias en precios o retornos y *I* la interacción entre dotaciones y retornos. Esta formulación permite dimensionar la brecha en términos absolutos, relativos y logarítmicos, así como identificar los mecanismos que amplifican o atenúan las desigualdades entre grupos.

### 3.4 Econometría espacial: modelo SLX

La metodología de econometría espacial, el modelo SLX incorpora explícitamente la dimensión territorial al análisis de la brecha salarial de género, para cuantificar no solo el efecto de los factores locales (efectos directos,  $\beta$ ) sino también los derrames de los municipios vecinos efectos de contexto o spillovers,  $\theta$ ) definidos por una matriz de pesos *W*. Aplicada a un panel municipal 2018 - 2024 permite identificar variables como formalidad educación o urbanización, tanto en el propio municipio como en su entorno, explican y reproducen la desigualdad, ofreciendo evidencia útil para políticas diferenciadas por territorio.

#### 3.4.1 Estimaciones previas y generación de variables

##### i) Depuración y transformación del ingreso

Se conservaron solamente ocupados con ingreso laboral positivo.

$$S = \{on: ing_{nit} > 0\}$$

##### ii) Winsorización para mitigar outliers

Se han calculado los percentiles  $p_1^{g,t}$  y  $p_{99}^{g,t}$  del ingreso nominaltra dentro de cada sexo  $g \in \{H, M\}$  y año  $t$  agregando todos los municipios para asegurar el tamaño muestral. (Se define el ingreso “winsorizado”) <sup>13</sup>

$$ing_{nit}^W = \min\{\max(ing_{nit}^{real}, p_1^{g,t}), p_{99}^{g,t}\}$$

Posteriormente se realiza una transformación logarítmica individual

$$l_{nit}^g = \ln(ing_{nit}^W), \quad g \in \{H, M\}$$

Se agrega la ponderación por municipio año y sexo.

Sea  $w_{nit}$  el factor de expansión de la ENOE. La media (ponderada) del Log-ingreso por municipio  $i$ , año  $t$  y sexo  $g$  es <sup>14</sup>:

$$\bar{y}_{it}^g = \frac{\sum_{n \in (i,t,g)} w_{nit} l_{nit}^g}{\sum_{n \in (i,t,g)} w_{nit}}$$

Generación de variable dependiente: brecha logarítmica municipal

$$brecha_{it} = \bar{y}_{it}^H - \bar{y}_{it}^M = \ln\left(\frac{G_{it}^H}{G_{it}^M}\right)$$

Donde la  $brecha_{it}$  es el logratio de los ingresos geométricos hombre/mujer en el municipio  $i$  y año  $t$ . En puntos porcentuales aproximados:

$$\% \Delta_{it} \approx 100 \cdot (e^{brecha_{it}} - 1)$$

es decir, si la  $brecha_{it} = 0.095 \Rightarrow$  ventaja masculina  $\approx 9.9\%$ <sup>15</sup>

<sup>13</sup> Si en algún  $g, t$  el  $n$  es pequeño, unos percentiles  $p_q^{g,t}$  como esperado. Esto estabiliza las colas sin perder señal.

<sup>14</sup> Equivalente:  $\bar{y}_{it}^g = \ln(G_{it}^g)$ , donde  $G_{it}^g$  es la media geométrica ponderada del ingreso en  $(i, t, g)$

<sup>15</sup> \* Bloques de p01/p99: cálculo de  $p_1^{g,t}$  y  $p_{99}^{g,t}$  y reemplazo para obtener  $ing^w$ .

\*  $\log\_ing$ : construcción de  $l_{nit}^g$

\*  $df\_H, df\_M$ : agregación ponderada por  $i - t - g$  para obtener  $\bar{y}_{it}^H - \bar{y}_{it}^M$ . Misma que es  $y_{it}$  en SAR/SLX

2) Se generaron brechas (gaps) en covariables para calcular covariable  $x$  y cada municipio  $i$  y año  $t$ , define la brecha como la diferencia hombre -mujer en el promedio ponderado por factores de expansión  $w$ .

$$\bar{x}_{it}^g = \frac{\sum_{n \in (i,t,g)} w_{nit} x_{nit}}{\sum_{n \in (i,t,g)} w_{nit}}, \quad g \in \{H, M\}$$

$$gap_{it}^x \equiv \bar{x}_{it}^H - \bar{x}_{it}^M$$

Esto hace que  $x_{it}$  (regresores) estén en la misma métrica (diferencia H-M) que la variable dependiente  $brecha_{it}$  (la diferencia en log-ingreso), mejorando la interpretación estructural del SAR/SLX.

Se formaron más “gaps” de las variables por ejemplo escolaridad

$$esc\_gap_{it} = \overline{años\_esc}_{it}^H - \overline{años\_esc}_{it}^M$$

(unidades: años, un valor =1 indica que, en promedio, los hombres del municipio  $i$  tienen unos años más de escolaridad que las mujeres en el año  $t$ ).

Formalidad (binaria):

Si  $formal_{nit} = \epsilon \{0,1\}$

$$\overline{formal}_{it}^g = \frac{\sum w_{nit} formal_{nit}}{\sum w_{nit}} \Rightarrow formal\_gap_{it} = \overline{formal}_{it}^H - \overline{formal}_{it}^M$$

(interpretación: diferencia en porcentaje de empleo formal entre hombres y mujeres)

Urbanidad (binaria):

$ur_{nit} = 1$  urbano, 0 rural

$$urbano\_gap_{it} = \overline{ur}_{it}^H - \overline{ur}_{it}^M$$

(distribución espacial de la oferta laboral por sexo; puede capturar ventajas urbanas en salarios y composición sectorial).

Edad

$$edad\_gap_{it} = \overline{edad}_{it}^H - \overline{edad}_{it}^M$$

Ramas (categóricas – proporciones)

Para cada rama  $c$  (agropecuario, industria y servicios). Se crean dummies  $d_{nit}^{(c)} \in \{0,1\}$  y calcular proporciones por cada sexo:

$$s_{it}^{(c)g} = \frac{\sum w_{nit} d_{nit}^{(c)}}{\sum w_{nit}}, \quad rama\_c\_gap_{it} = s_{it}^{(c)H} - s_{it}^{(c)M}$$

Por lo tanto, tenemos en cada categoría:

$$rama\_agro\_gap_{it}, rama\_inds\_gap_{it} \text{ y } rama\_serv\_gap_{it}$$

Interpretaciones esperadas (intuición econométrica)

- $\beta_{esc} > 0$ : municipios donde los hombres aventajan en escolaridad tienden a presentar mayor brecha salarial a su favor
- $\beta_{formal} > 0$ : mayor proporción masculina en formalidad (con prima salarial) ampliaría la brecha.
- $\beta_{urb} > 0$ : ambiguo; si las mujeres están relativamente mas en zonas urbanas (mejores retornos) podría reducir la brecha (coef negativo).
- $\beta_{edad} > 0$ : si los hombres concentran más experiencia efectiva; podría revertirse si hay penalizaciones por edad en ocupaciones feminizadas.
- $\beta_{rama} > 0$ : signos dependen de las primas sectoriales locales (industria suele pagar mas que agropecuario y comercio).

Esto se hace para definir los  $X_{it}$  de SLX/SAR:

$$brecha_{it} = \rho Wbrecha_{it} + X_{it}\beta + WX_{it}\theta + u_{it}$$

Donde  $X_{it}$  es el vector de “gaps”.

En seguida se realiza la selección parsimoniosa de predictores, buscando un conjunto mínimo de “gaps”  $X_{it}$  informativos y con baja colinealidad mismas que se trabajan a nivel municipio y año.

1. Estandarización de cada gap para comparación

$$z_{it}^{(k)} = \frac{x_{it}^{(k)} - \overline{x^{(k)}}}{s^{(k)}}, \quad k = 1, \dots, K$$

2. Matriz de correlaciones:

$$R = [r_{ab}] = \text{Corr!}(z^{(a)}, z^{(b)}) = \frac{1}{N-1} Z'Z,$$

Donde  $Z$  apila los  $z^{(k)}$  y  $N$  es el número de celdas  $(i, t)$ .

3. VIF (inflación de varianza) para cada predictor  $x_k$  via la regresión auxiliar sobre  $x_{-k}$

$$R_k^2 = R^2(x_k \sim X_{-k}), \quad \text{VIF}_k = \frac{1}{1 - R_k^2}.$$

De acuerdo con el criterio:  $\text{VIF}_k < 5$  (estricto)  $0 < 10$  tolerable.

índice de condición (estabilidad global):

$$\kappa = \sqrt{\frac{\lambda_{\max}(X'X)}{\lambda_{\min}(X'X)}}.$$

De manera que el conjunto parsimonioso y robusto queda de la siguiente manera:

$$X_{it} = (\text{esc\_gap}_{it}, \text{formal\_gap}_{it}, \text{urbano\_gap}_{it}, \text{edad\_gap}_{it}, \text{rama\_industria\_gap}_{it}),$$

ecuación objetivo:

$$\frac{\text{brecha}_{it}}{Y_{it}} = \rho \sum_j w_{ij} \text{brecha}_{jt} + \sum_{\text{kvars\_gap\_final}} \beta_k x_{it}^{(k)} + \sum_{\text{kvars\_gap\_final}} \theta_k (WX)_{it}^{(k)} + u_{it}.$$

Estructura espacial: matriz de pesos y rezagos de X (WX)<sup>16</sup>

Se alinearon los municipios entre shapefiles (polígonos) y el panel (i,t) para que la fila/columna  $i$  de  $W$  corresponda exactamente al municipio  $i$  en los datos.

Sea  $S_{shp} = \{i: CVEGEO_i \text{ en shapefile}\}$  y  $S_{dat}(t) = \{i:$

$CVEGEO_i$  en la base para el año  $t\}$

- Intersección común en todo el panel:

$$\mathcal{M}^* \equiv \bigcap_t (S_{shp} \cap S_{dat}(t)).$$

- $N = |\mathcal{M}^*|$  Municipios estables

1. Un vector de  $IDS\{CVEGEO_i\}_{i=1}^N$  con correspondencia perfecta datos-geometría
2. Un orden fijo  $i = 1, \dots, N$  para construir la matriz de adyacencia  $A = [a_{ij}]$  y, luego la matriz de pesos  $W = [w_{ij}]$  fija estandarizada.

$$w_{ij} = \frac{a_{ij}}{\sum_j a_{ij}}, \quad w_{ii} = 0.$$

3. Con ese  $W$ , los rezagos espaciales de cualquier variable  $x^{(k)}$

$$(WX)_{it}^{(k)} = \sum_{j=1}^N w_{ij} x_{jt}^{(k)}.$$

De la variable de los municipios vecinos, se genera un promedio ponderado con los pesos, de la matriz donde  $(WX)_{it}$  mide la influencia promedio de los vecinos.

### 3.4.2 Modelos SLX (spatial Lag o X)

Ecuación (panel municipio-año)

---

<sup>16</sup> El modelo SLX se inscribe en la tradición de la econometría espacial desarrollada por Anselin (1988) y sistematizada para paneles por Elhorst (2003), y permite identificar si las condiciones de los municipios vecinos ejercen efectos indirectos sobre los niveles y la dinámica de la brecha salarial de género.

$$\underbrace{brecha_{it}}_{Y_{it}} = \alpha_i + \delta_t + \underbrace{X'_{it}\beta}_{\text{efecto local}} + \underbrace{(WX_t)'_i\theta}_{\text{efectos de vecinos}} + \varepsilon_{it}$$

- $\alpha_i$  son “cosntantes” por municipio: todo lo que no cambia en el tiempo y hace distintos a los municipios (cultura laboral, patrón histórico de especialización, instituciones locales).
- $\delta_t$  son “contantes” por años: shocks nacionales omunes (SMG federales, pandemia, ciclo macro).
- $X_{it}$  es el vector de gaos locales (educación formalidad, urbanización, edad, ramas).
- $(WX_t)_i$  es el promedio ponderado en los vecinos de esas mismas variables. La matriz  $W$  contiene peses espaciales (p.j. contigüidad tipo queen<sup>17</sup>, fila-estandarizada), de modelo que  $\sum_j w_{ij} = 1$  y  $w_{ii} = 0$ .

La brecha en  $i$  es la suma de su “nivel propio” ( $\alpha_i$ ), lo que pasa en el país ese año  $\delta_t$ , los determinantes locales  $X_{it}$  y el entorno de los vecinos  $(WX_t)_i$

- $\beta_k$  mide cuanto cambia la brecha local si el gap  $k$  en el propio municipio sube una unidad, manteniendo lo demás fijo.
- $\theta_k$  mide cuanto cambia la brecha local si el promedio en los vecinos del gap  $k$  sube una unidad.

Como no metemos  $WY$  (la brecha de los vecinos) en la derecha, no hay retroalimentaciones ni efectos cascada; por eso, en SLX, los, “impactos” coinciden directamente con  $\beta$  (local) y  $\theta$  (vecinos).

### 3.4.2.1 Efectos fijos

Suponemos que hay factores inobservables propios del municipio que no vemos y podrían estar correlacionados con  $X$ . si no los controlamos, sesgan los coeficientes. Con efectos fijos ( $\alpha_i$ )

---

<sup>17</sup> En los modelos espaciales, la forma en que define la matriz de pesos  $W$  resulta clave. Entre más empleadas se encuentra la denominada *queen*, que establece vecindad cuando dos unidades comparten limite o vértice, y el esquema de  $k$ -vecinos más cercanos ( $k$ -NN), que asigna como adyacentes a las  $k$  áreas más próximas en términos de distancia. La primera capta relaciones amplias de contigüidad, mientras que la segunda asegura un numero uniforme de conexiones para cada unidad, lo que hace útil en territorios de tamaños desiguales.

retiramos esa heterogeneidad invariante;  $\delta_t$  limpia shocks comunes por año. Formalmente, estimamos el modelo “dentro” (within).<sup>18</sup>

### 3.4.2.2 Revisión espacial Ex-post

Tras estimar, miramos si en los residuos aún queda patrón espacial. Calculamos el *Moran's I* por año:

$$I_{Moran}(\hat{e}_t) = \frac{n}{1'W1} \frac{\hat{e}_t'W\hat{e}_t}{\hat{e}_t'\hat{e}_t}$$

- Si no es significativo, entendemos que WX ya absorbió la dependencia espacial relevante.
- Si sí queda dependencia, entonces el fenómeno puede requerirlo un SAR/SEM (esto es, modelar también la interacción espacial de la propia brecha o del error).

Reporte de resultados del índice de Moran

- Coeficientes  $\beta$  y  $\theta$  con EE robustos y su interpretación en puntos porcentuales (usando  $100 \cdot (e^\Delta - 1)$ ).
- F-test de efectos fijos, prueba de Hausman,  $R^2$  within.
- Moran's I de los residuos por año.
- Sensibilidades: otra W (k-NN, distancia), otro set de predictores y versión sin winsorización.

Modelo SAR (Spatial Autoregressive), contraste al SLX

Plantea que la brecha de un municipio no solo responde a sus covariables sino también a la brecha de sus vecinos:

$$\text{brecha}_{it} = \rho (W_t \text{brecha}_t)_i + X'_{it}\beta + u_{it}, \quad u_{it} \sim \text{i.i.d.}$$

- $(W_t \text{brecha}_t)_i = \sum_j w_{ij} \text{brecha}_{jt}$  es el promedio ponderado de la brecha en los municipios cotiguos a  $i$  (con  $W$  fila-estandarizada y  $w_{ij} = 0$ ).

---

<sup>18</sup> Estimamos con FX y tratamos con RE con Hausman. Si Hausman rechaza, nos quedamos con FX

- $\rho$  mide la dependencia espacial en la propia dependiente: si  $\rho > 0$ , un aumento de la brecha en los vecinos tiende a empujar hacia arriba la brecha local; si  $\rho \sim 0$ , no hay propagación;  $\rho < 0$  implicaría efectos compensatorios (menos usual).
- $X_{it}\beta$  son los efectos locales de los “gaps” (educación, formalidad, etc.).
- Aquí no se incluye  $WX$ ; el foco está en spillovers de la propia brecha, no en los de las covariables (mismo que captura SLX/SDM).

Propagación del shock

$$brecha_t = (I + \rho W_t)^{-1} X_t \beta + (I - \rho W_t)^{-1} u_t.$$

El factor  $(I + \rho W_t)^{-1}$  es un multiplicador espacial: un cambio inicial en un municipio rebota hacia sus vecinos, a los vecinos de los vecinos, etc. Mientras mayor  $\rho$  (y conectividad de  $W$ ), más largo es el eco espacial del Shock.<sup>19 20</sup>

Lectura

En SAR, los coeficientes  $\beta$  no son el efecto final 1 a-1 porque se multiplican por  $(I - \rho W)^{-1}$ .

Se reportan impactos (LaSage y Pace):

- Directo: Promedio del bloque diagonal de  $(I - \rho W)^{-1} \beta$  (efecto sobre el propio municipio).
- Indirecto (spillover): promedio de los elementos fuera de la diagonal (lo que se transmite a/desde vecino).
- Total: directo + indirecto.

---

<sup>19</sup> SLX pregunta: “¿las condiciones de mis vecinos (sus gaps) afectan mi brecha?”

SAR pregunta: “¿la brecha de mis vecinos arrastra mi brecha?”

Si tras estimar SLX quedan residuos con autocorrelación espacial (Moran’s I significativa), el SAR es un chequeo de robustez natural: modela directamente la interacción de la propia  $Y$ .

<sup>20</sup> El término  $Wbrecha$  está endógeno (contiene  $Y$  en el lado derecho). Por eso:

- En corte transversal (por año): usar Máxima Verosimilitud o IV/GMM; en R,
- En panel con efectos fijos: usar estimadores específicos, que tratan simultáneamente  $\alpha_i$  y  $\delta_t$ .

### 3.4.3 Implementación operativa

A partir del panel municipal 2018-2024 de microdatos de la ENOE<sup>21</sup> se realiza la depuración, winsorización y log, se conserva a ocupados con ingreso positivo. Para cada observación  $n$  (persona), sexo  $g \in \{H, M\}$ , municipio  $i$  y año  $t$ :

Winsor con percentiles  $p_1^{g,t}, p_{99}^{g,t}$ :

$$\text{ing}_{nigt}^W = \min\{\max(\text{ing}_{nigt}^{\text{real}}, p_1^{g,t}), p_{99}^{g,t}\}$$

Log individual:

$$l_{nigt} = \ln(\text{ing}_{nigt}^W)$$

Posteriormente se agrega el ponderador de la ENOE por sexo municipio y año, con factores de expansión  $w_{nigt}$ :

$$\bar{y}_{it}^g = \frac{\sum_{n \in (i,t,g)} w_{nigt} l_{nigt}}{\sum_{n \in (i,t,g)} w_{nigt}} = \ln G_{it}^g$$

Donde  $G_{it}^g$  es la media geométrica ponderada del ingreso.

Brecha logarítmica de ingresos:

$$\text{brecha}_{it} = \bar{y}_{it}^H - \bar{y}_{it}^M = \ln\left(\frac{G_{it}^H}{G_{it}^M}\right)$$

Para cualquier covariable  $x$  (años de escolaridad, empleo formal, urbano, edad, ramas económicas), el gap municipal es:

$$\text{gap}_{it}^{(x)} = \bar{y}_{it}^H - \bar{y}_{it}^M, \quad \bar{x}_{it}^g = \frac{\sum w_{nigt} x_{nigt}}{\sum w_{nigt}}$$

Matriz espacial anual  $W_t$  y rezagos  $W_t X_t$

De acuerdo con el marco geoestadístico INEGI (polígonos municipales), se construye en base a CVEGEO<sup>22</sup> y se cruza con los IDs de la base para quedarse con la intersección estable en todo el panel.

<sup>21</sup> Si bien el título de la investigación de acota al período 2018-2023, durante el desarrollo del estudio se incorporaron las últimas actualizaciones disponibles de la ENOE correspondientes a 2024, con el fin de aprovechar la base de panel completa del sexenio completo de igual manera ayuda a fortalecer la robustez de los resultados.

<sup>22</sup> Clave Única de Áreas Geoestadísticas, permite ubicar geográficamente la información estadística y censal, donde los registros estén correctamente vinculados a su espacio geográfico.

Estandarización por filas: <sup>23</sup>

$$W_t = D_t^{-1}A_t, \quad (D_t)_{ii} = \sum_j a_{ij,t}, \quad w_{ii,t} = 0, \quad \sum_j w_{ij,t} = 1.$$

Estimación SLX con efectos fijos de dos vías

$$\text{brecha}_{it} = \alpha_i + \delta_t + X'_{it}\beta + (W_t X_t)'_i \theta + \varepsilon_{it}$$

- $\alpha_i$ : heterogeneidad invariante municipal (capturada con within).
- $\delta_t$ : shocks anuales comunes (pandemia, SMG, ciclo).
- $\beta$ : efectos locales;  $\theta$ : efectos de contexto vecinal.

Esquema within:

$$y_{it}^* = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot}, \quad X_{it}^* \text{ analogo};$$

OLS sobre  $(y^*, X^*)$  entrega lo FE.

Decisión FE y RE (Hausman)

Se estima el modelo RE, y se corre:

$$H = (\widehat{\beta}_{RE} - \widehat{\beta}_{FE})^\top [\text{Var}(\widehat{\beta}_{RE}) - \text{Var}(\widehat{\beta}_{FE})]^{-1} (\widehat{\beta}_{RE} - \widehat{\beta}_{FE})$$

- P-valor < 0.05; se prefiere FE.
- EE robusto: tipo HC con clúster por municipio o Driscoll-Kraay.

SLX anual y Moran de residuos

Para cada año  $t$  se estima lmSLX (corte transversal):

$$I_{\text{Moran}}(\widehat{e}_t) = \frac{n}{1'W_t1} \frac{\widehat{e}_t^\top W_t \widehat{e}_t}{\widehat{e}_t^\top \widehat{e}_t}.$$

- Tras el SLX: no hay autocorrelación residual (p-valores altos), señal de que  $W_t X_t$  ya absorbió la dependencia espacial relevante.

Mediante este análisis encontramos importantes efectos como el de la formalidad (gap H-M) en la brecha salarial

<sup>23</sup> (zero.police=TRUE evita NAs cuando un municipio caree de vecinos; si existieran “islas” se suele complementar con k-NN).

$$\hat{\beta}_{formal\_gap} \approx 0.56$$

Un aumento de 10 p.p. en el gap de formalidad masculina local implica:

$$\Delta \ln(\text{brecha}) \approx 0.56 \times 0.10 = 0.056 \Rightarrow \% \Delta \text{brecha} \approx 100 \cdot (e^{0.056} - 1) \approx +5.7\%$$

Donde los hombres superan a las mujeres en empleo formal, la brecha salarial es significativamente mayor.

De igual manera efectos en la escolaridad (gap en años).

$$\hat{\beta}_{formal\_gap} \approx 0.56$$

donde +1 año de escolaridad masculina sobre femenina:

$$\% \Delta \text{brecha} \approx 100 \cdot (e^{0.048} - 1) \approx +4.9\%$$

También encontramos un rezago robusto

$$\hat{\theta}_{WX_{formal}} > 0.$$

En este, existe difusión territorial de la desigualdad de género en empleo formal: un entorno con alta ventaja masculina en formalidad empuja al alza la brecha salarial del municipio aun controlando por sus propios gaps.

Mientras que el índice de Moran en los residuos (post-SLX) resulta no significativo año con año → la especificación captura la estructura espacial son requerir dependencia en  $Y$  (SAR) ni el error (SEM). Esto valida usar SLX como modelo principal y dejar SAR/SEM/SDM como robustez.

### 3.5 Consideraciones de identificación y limitación

Los resultados econométricos de este trabajo deben interpretarse como relaciones condicionales y no como efectos causales en sentido estricto. La identificación se basa en el supuesto de que, una vez condicionados los modelos por el conjunto de covariables observadas (capital humano, características ocupacionales, rasgos del hogar y del territorio), el término de error es ortogonal a dichas variables. En la práctica, es posible que persistan factores no observados como

habilidades específicas, actitudes hacia el riesgo, características del empleador o información imperfecta que estén correlacionados tanto con el salario como con el género, lo que introduce un posible sesgo de variables omitidas.

En el modelo de corrección por selección tipo Heckman, la identificación descansa además en la existencia de al menos una variable que afecte la probabilidad de participar en el mercado laboral con salario observado, pero que no influya directamente en el nivel del salario una vez condicionadas las demás covariables. Aunque se incorporan variables relacionadas con la estructura del hogar y las responsabilidades de cuidado, es posible que estos instrumentos no sean perfectamente excluyentes, por lo que los resultados deben leerse como una aproximación al sesgo de selección, más que como su corrección completa.

En el modelo espacial SLX, se asume que la matriz de pesos  $W$  es exógena y fija en el tiempo y que los rezagos espaciales de las covariables capturan adecuadamente los efectos de vecindad. No se modela la endogeneidad potencial de la brecha salarial municipal ni la posible retroalimentación entre municipios, por lo que los spillovers estimados deben interpretarse como correlaciones espaciales condicionadas, y no como efectos causales de política.

En conjunto, estos supuestos implican que los resultados ofrecen una caracterización robusta de los patrones de desigualdad salarial de género y sus correlatos observables, pero no permiten afirmar que las diferencias observadas sean efectos causales puros de cada covariable.

## CAPÍTULO IV. Resultados

### 4.1 Estimación Oaxaca Blinder

Esta sección organiza los resultados descriptivos y la descomposición salarial por género para tres cortes comparables del mercado laboral mexicano: 2019 (pre-choque sanitario), 2021 (reacomodo postpandemia) y 2023 (fase de normalización). La estimación se ejecutó en Stata mediante la metodología de Oaxaca Blinder, siguiendo la implementación canónica de (Jann, 2008), lo que permite separar el diferencial salarial en una parte asociada a dotaciones observables y otra vinculada a retornos diferenciales / factores no observados. El objetivo aquí es doble: (i) dimensionar la brecha en términos absolutos, relativos y logarítmicos; y (ii) anticipar que fracción del diferencial obedece a composiciones frente a lo que persiste como injustificado desde los datos.

Tabla 4.1 Evolución salarial comparada

<b>Concepto</b>	<b>2019</b>	<b>2021</b>	<b>2023</b>
Salario hombres	\$ 5,162.06	\$ 5,764.16	\$6,649.80
Salario mujeres	\$ 3,848.39	\$ 4,493.00	\$5,060.24
Diferencia absoluta	\$ 1,313.67	\$ 1,271.17	\$1,589.56
Brecha en log-salario	0.128	0.108	0.119
<i>Brecha porcentual</i>	25.45%	22.05%	23.90%
Parte explicada	0.0444	0.03475	0.056
Parte no explicada	0.305	0.2516	0.2529

Fuente: Creación propia con datos de ENOE 2019, 2021 y 2023.

La Tabla 4.1 resume la trayectoria de los promedios salariales por sexo y la brecha correspondiente. En 2019, el ingreso medio de los hombres fue de \$5,162.06, mientras que el de las mujeres alcanzó \$3,848.39. La diferencia absoluta fue de \$1,313.67, con una brecha porcentual de 25.45% y una brecha en log-salario de 0.128. Para 2021, ambos promedios aumentaron a \$5,764.16 y \$4,493.00. Con ello, el diferencial absoluto bajó a \$1,271.17. La

brecha relativa se atenúa a 22.05% (log 0.108). En 2023, los promedios vuelven a crecer (\$6,649.80 y \$5,060.24), pero el diferencial en pesos sube a \$1,589.56, con una brecha porcentual de 23.90% y una brecha en log de 0.119.

Tabla 4.2 Descomposición doble Oaxaca-Blinder: Factores Explicativos y No Explicativos de las Diferencias Salariales (2019, 2021 y 2023)

Característica	2019		2021		2023	
	Explicada	No Explicada	Explicada	No Explicada	Explicada	No Explicada
<i>Rural</i>	-0.0044***	-0.0034*	-0.0029***	0.0149***	-0.0026***	0.0092***
<i>No sabe leer/escribir</i>	-0.0004***	-0.0005	-0.0002***	0.0015***	-0.0004***	0.0001
<i>Estado Civil</i>	0.0092***	0.0641***	0.0057***	0.0534***	0.0079***	0.0812***
<i>Subocupado</i>	-0.0033***	0.0008	-0.0016***	-0.0049***	-0.0003	0.0054***
<i>Zona A</i>	-0.0011**	-0.0044***	0.0014***	-0.0055***	0.0001	-0.0058***
<i>Act. Primarias</i>	0.0016	-0.0104***	-0.0129***	-0.0088***	0.0038***	-0.0157***
<i>Act. Secundarias</i>	0.0429***	0.0660***	0.0149***	0.0703***	0.0403***	0.0185**
<i>Act. Terciarias</i>	-0.0650***	-0.016	-0.0119***	0.1078***	-0.0477***	-0.0126
<i>Dos salarios</i>	-0.0047***	-0.1473***	0.0058***	-0.1306***	0.0241***	-0.1045***
<i>Mas de tres salarios</i>	0.0846***	-0.0481***	0.0704***	-0.0218***	0.0572***	-0.0124***
<i>Servicios medicos</i>	0.0019**	-0.1671***	-0.0024***	-0.0993***	0.0010**	-0.0842***
<i>Menor a 14 años</i>	-0.0031***	-0.0014***	-0.0012***	-0.0008***	-0.0020***	-0.0005***
<i>14 - 24 años</i>	0.0034***	-0.0181***	0.0017***	-0.0088***	0.0003**	-0.0117***
<i>24 - 34 años</i>	-0.0017**	-0.0057	-0.0014***	-0.0152***	0	-0.0071***
<i>34 - 44 años</i>	0.0007	-0.0055	-0.0074***	-0.0070**	-0.0042***	-0.0079***
<i>44 - 54 años</i>	-0.0060***	-0.0142***	-0.0039***	-0.0123***	-0.0051***	-0.0079***
<i>54 - 64 años</i>	-0.0010***	0.0196***	0.0012***	-0.0048***	0.0002	-0.0026*
<i>Sin estudios</i>	-0.0002*	-0.0008	-0.0004*	-0.0005	-0.0003***	-0.0008*
<i>Preescolar</i>	0	0	0	0	0	0
<i>Primaria</i>	-0.0034***	0.0071**	-0.0039***	0.0078***	-0.0044***	0.0047***
<i>Secundaria</i>	-0.0004	0.0405***	-0.0043***	0.0231***	-0.0029***	0.0065***
<i>Preparatoria</i>	-0.0032***	0.0185***	-0.0026***	0.002	-0.0019***	0.0022
<i>Profesional</i>	-0.0012**	0.0243***	-0.0068***	-0.0251***	-0.0051***	-0.0218***
<i>Posgrado</i>	-0.0007***	0.0011*	-0.0023***	-0.0027***	-0.0019***	-0.0030***

Fuente: Creación propia con datos de ENOE 2019 2021 y 2023 INEGI. Nivel de significancia \* p < 0.05; \*\* p < 0.01; \*\*\*p < 0.001.

En conjunto, los ingresos aumentan para ambos grupos a lo largo del período. Se observa un cierre temporal del diferencial en 2021 y una ligera reapertura en 2023, lo que sugiere una mejora general de los salarios reales sin eliminar el sesgo por sexo.<sup>24</sup>

Previo al análisis del modelo Oaxaca-Blinder, el Cuadro 4.1 permite fijar la idea de que la porción explicada por diferencias observables es limitada, mientras que el tramo no explicado domina la brecha en los tres años. En términos de log-salario, la parte explicada ronda 0.0444 en 2019, 0.03475 en 2021 y 0.056 en 2023, lo que equivale aproximadamente a 13%, 12% y 18% del diferencial total, respectivamente. En consecuencia, más del 80% de la brecha se mantiene como no explicada a lo largo del período. Esta estabilidad del componente residual, pese a la mejora salarial promedio, constituye un indicador robusto de retornos asimétricos<sup>25</sup> y/o mecanismos institucionales que continúan penalizando a las mujeres.

La tabla 4.2 profundiza la lectura al desagregar la descomposición doble (explicada / no explicada) por grupos de variables para 2019, 2021 y 2023. A grandes rasgos, se observan cuatro patrones consistentes:

(i) Estado Civil

El estado civil empuja la brecha por dos canales: composición (parte explicada positiva) y retornos (parte no explicada positiva). Esto significa que las diferencias en la distribución por estado conyugal ya aplican el diferencial y, además, al igual estado, los retornos salariales favorecen a los hombres. El efecto es significativo y persistente en los tres cortes y se refuerza en 2023.

(ii) Educación

En niveles básicos y medio superior (primaria, secundaria, preparatoria) la parte explicada es negativa: las diferencias de capital humano amortiguan la brecha, en línea con la mayor acumulación educativa femenina en varios tramos; sin embargo, los retornos no son homogéneos. En profesional y posgrado se observa un cuadro matizado: la composición tiende

---

<sup>24</sup> La brecha el log-salario es la escala natural para la descomposición econométrica, pues interpreta variaciones como semi-elasticidades y estabiliza la asimetría típica de los ingresos.

<sup>25</sup> Los retornos asimétricos aluden a que, ante iguales niveles de educación o experiencia, las mujeres reciben menores beneficios salariales que los hombres.

a cerrar el diferencial (explicada negativa), mientras que el componente no explicado es positivo en 2019 (y en algunos casos cambia de signo en 2021–2023). La lectura económica es clara: a credenciales similares, la recompensa salarial no es la misma en todos los años, lo que apunta a penalizaciones de retorno en la parte alta del capital humano.

### (iii) Estructura Sectorial

En actividades secundarias (industria) tanto la composición como los retornos suelen ensanchar la brecha (efectos positivos y significativos en 2019 y 2021; en 2023 persiste el sesgo, aunque el tramo no explicado se modera). Por el contrario, en actividades terciarias la parte explicada aparece con signo negativo (la composición reduce la brecha), pero en 2021 emerge un componente no explicado positivo y elevado, indicio de que, en un año de recomposición postchoque, a la igual inserción sectorial la retribución favoreció al trabajador varón. Este contraste sectorial sugiere que donde se emplea la fuerza laboral femenina y como se remunera dentro del sector no son cuestiones equivalentes.

### (iv) Ámbito territorial y condiciones laborales.

En ruralidad y zona (clasificación espacial), la composición tiende a mitigar el diferencial (explicada negativa), pero se observan no explicados positivos en 2021–2023, señalando que a igual localización persisten retornos diferenciales por sexo. Respecto a variables de umbral salarial (dos y más de tres salarios mínimos) y servicios médicos, aparecen no explicados negativos en varios cortes, lo que denota heterogeneidad en la estructura de primas y beneficios: no toda “cúpula de ingreso” se traduce automáticamente en mayor brecha, ni todo acceso a servicios se remunera con el mismo sesgo; la dirección del efecto depende de cómo se paga y a quién alcanza, año con año. En conjunto, la evidencia reafirma que las asimetrías de retorno (no explicadas) son las que definen la magnitud final del diferencial.

El patrón interanual (brecha elevada en 2019, disminución parcial en 2021 y reapertura en 2023), posee dos mecanismos coexistentes y coherentes: piso pegajoso en la base de la distribución, donde la recomposición del empleo femenino tras el choque sanitario no traduce en movilidad sostenida hacia tramos mejor pagados y “techo de cristal” en la parte alta de la distribución donde la subrepresentación femenina y los retornos asimétricos en calificaciones altas y sectores con primas elevadas reactivan el diferencial una vez que la economía retoma ritmo.

Tabla 4.3 Análisis de regresión en las variables socioeconómicas de hombres y mujeres (2019, 2021 y 2023)

Variable	2019		2021		2023	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
<i>Rural</i>	-0.058*** (0.002)	-0.046*** (0.004)	-0.041*** (0.002)	-0.082*** (0.003)	-0.035*** (0.002)	-0.060*** (0.003)
<i>No sabe leer/escribir</i>	-0.163*** (0.015)	-0.143*** (0.029)	-0.046*** (0.013)	-0.143*** (0.023)	-0.220*** (0.015)	-0.228*** (0.020)
<i>Estado civil</i>	0.107*** (0.003)	-0.001 (0.004)	0.082*** (0.002)	-0.011** (0.003)	0.127*** (0.002)	-0.015*** (0.003)
<i>Subocupado</i>	-0.170*** (0.005)	-0.187*** (0.009)	-0.208*** (0.004)	-0.169*** (0.005)	-0.207*** (0.004)	-0.276*** (0.006)
<i>Zona A</i>	0.276*** (0.005)	0.322*** (0.006)	0.205*** (0.003)	0.273*** (0.005)	0.269*** (0.003)	0.338*** (0.004)
<i>Act. primaria</i>	-0.064 (0.039)	0.200*** (0.052)	-0.219*** (0.022)	0.038 (0.024)	-0.078*** (0.023)	0.231*** (0.021)
<i>Act. secundaria</i>	0.365*** (0.038)	0.095* (0.050)	0.189*** (0.021)	-0.108*** (0.022)	0.238*** (0.022)	0.127*** (0.020)
<i>Act. Terciaria</i>	0.296*** (0.038)	0.317*** (0.050)	0.121*** (0.021)	-0.036 (0.022)	0.171*** (0.022)	0.197*** (0.020)
<i>Dos salarios</i>	-0.013*** (0.003)	0.365*** (0.004)	0.121*** (0.002)	0.407*** (0.003)	0.240*** (0.002)	0.476*** (0.003)
<i>Más de tres salarios</i>	0.924*** (0.004)	1.209*** (0.006)	1.098*** (0.003)	1.273*** (0.005)	1.236*** (0.004)	1.403*** (0.006)
<i>Servicios médicos</i>	0.105*** (0.002)	0.465*** (0.004)	0.185*** (0.002)	0.400*** (0.003)	0.173*** (0.002)	0.360*** (0.003)
<i>Menor a 14 años</i>	-0.817*** (0.023)	-0.464*** (0.053)	-0.622*** (0.022)	-0.435*** (0.033)	-0.744*** (0.020)	-0.609*** (0.039)
<i>14 - 24 años</i>	0.084*** (0.010)	0.201*** (0.015)	0.052*** (0.007)	0.111*** (0.011)	-0.003 (0.006)	0.077*** (0.010)
<i>24 - 34 años</i>	0.260*** (0.009)	0.284*** (0.014)	0.214*** (0.007)	0.277*** (0.011)	0.176*** (0.006)	0.206*** (0.010)
<i>34 - 44 años</i>	0.267*** (0.009)	0.290*** (0.014)	0.241*** (0.006)	0.268*** (0.010)	0.184*** (0.006)	0.216*** (0.010)
<i>44 - 54 años</i>	0.260*** (0.009)	0.327*** (0.014)	0.231*** (0.007)	0.290*** (0.010)	0.202*** (0.006)	0.238*** (0.009)
<i>54 - 64 años</i>	0.175*** (0.010)	0.006 (0.015)	0.145*** (0.007)	0.188*** (0.011)	0.130*** (0.006)	0.151*** (0.010)

Fuente: Creación propia a través de datos ENOE 2019, 2021 y 2023 del INEGI. Dos columnas (H|M) con coeficientes y EE. Nota: modelos OLS sobre  $\ln(w)$ ; dummies categorías con base explicada; pesos muestrales ENOE; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

Tabla 4.3.1 (continuación) Análisis de regresión en las variables socioeconómicas de hombres y mujeres (2019, 2021 y 2023)

Variable	2019		2021		2023	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
<i>Sin estudios</i>	-0.182*** (0.017)	-0.137*** (0.027)	-0.278*** (0.015)	-0.248*** (0.022)	-0.173*** (0.015)	-0.131*** (0.019)
<i>Primaria</i>	-0.098*** (0.008)	-0.138*** (0.010)	-0.129*** (0.007)	-0.178*** (0.008)	-0.174*** (0.007)	-0.207*** (0.007)
<i>Secundaria</i>	-0.026*** (0.007)	-0.157*** (0.009)	-0.085*** (0.006)	-0.167*** (0.007)	-0.121*** (0.006)	-0.145*** (0.007)
<i>Preparatoria</i>	0.006 (0.007)	-0.071*** (0.009)	-0.056*** (0.006)	-0.065*** (0.007)	-0.055*** (0.006)	-0.065*** (0.007)
<i>Profesional</i>	0.085*** (0.008)	-0.029** (0.009)	0.068*** (0.007)	0.171*** (0.007)	0.069*** (0.006)	0.158*** (0.007)
<i>Posgrado</i>	0.191*** (0.013)	0.143*** (0.015)	0.175*** (0.010)	0.282*** (0.010)	0.164*** (0.009)	0.282*** (0.010)
<i>Constante</i>	7.954*** (0.040)	7.445*** (0.053)	8.269*** (0.023)	7.952*** (0.025)	8.407*** (0.024)	7.984*** (0.023)

Fuente: Creación propia a través de datos ENOE 2019, 2021 y 2023 del INEGI. Dos columnas (H/M) con coeficientes y EE. Nota: modelos OLS sobre  $\ln(w)$ ; dummies categorías con base explicada; pesos muestras ENOE; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

El desglose de la tabla 4.2 refina la interpretación: el estado civil y la composición sectorial empujan el diferencial en el sentido previsto (ampliándolo), mientras que la educación —por su efecto de composición— lo amortigua; no obstante, ello ocurre a costa de que las mujeres obtengan retornos más bajos en los tramos altos en varios años. En suma, la convergencia de dotaciones no es suficiente; lo decisivo es cómo el mercado remunera esas dotaciones según el género.

La Tabla 4.3 presenta los resultados del modelo lineal en log-salario estimado por separado para hombres y mujeres en los tres cortes anuales respectivos. Las covariables incluyen variables ya previamente expuestas. En todos los casos se reportan errores estándar y significancia, lo que permite comparar retornos entre géneros y a través del tiempo. Estos coeficientes son la base empírica para contrastar la hipótesis de retornos asimétricos por género (misma dotación, precio distinto).

El cuadro 4.3 y 4.3.1 exponen en general resultados transversales (retornos y asimetrías) a través de las dimensiones:

(i) Territorio y condiciones del puesto.

Ruralidad. En los tres cortes, residir en ámbito rural reduce el log-salario, pero la penalización es sistemáticamente mayor en mujeres. En 2023 el descuento ronda ~6% para ellas y ~3.5% para ellos de acuerdo con las estimaciones puntuales del modelo. Esto sugiere que idéntica localización geográfica se traduce en retornos más bajos para las trabajadoras, quizá por menos acceso a puestos formales o a ocupaciones con primas en el retorno rural.

Subocupación. Estar subocupado implica una pérdida salarial importante para ambos sexos; sin embargo, la penalización más severa es en mujeres y se intensifica en 2023 ( $\approx -27.5\%$  en ellas vs.  $\approx -20.7\%$  en ellos). Este diferencial es coherente con rutas ocupacionales más inestables para la fuerza laboral femenina, así como con una menor capacidad de negociación en jornadas y asignaciones.

Zona A. (alta remuneración relativa). Las dummies territoriales tipo “Zona A” se asocian con primas salariales significativas en ambos géneros, ligeramente mayores para mujeres; esto indica que, cuando el puesto se ubica en espacio de mayor dinamismo (industria o metropolitano), ellas también capturan de esa prima.

(ii) Estructura Sectorial

Actividades primarias. En 2023 se observa una penalización para los varones (coeficiente negativo y significativo), mientras que para mujeres aparece un premio relativo. Este contraste —aparente “prima” femenina— suele revelar selección hacia nichos ocupacionales específicos dentro del sector (p. ej., empleos formales/organizados o de mayor calificación) y no una ventaja sectorial generalizada para ellas. Requiere cautela: el efecto es relativo al sector base y puede estar mediado por composición interna de ocupaciones agrícolas/pecuarias.

Actividades secundarias. La industria otorga primas claras para ambos sexos en los tres años, con retornos superiores para los varones. Ello sugiere que, aunque la puerta industrial mejora salarios de todos, ellos capturan una parte mayor de la renta sectorial, consistente con posiciones técnicas/jerárquicas donde la brecha reaparece vía salarios y bonos.

Actividades terciarias. Los servicios muestran premios positivos para ambos en 2019 y 2023; en 2021 la señal femenina se atenúa. Dado que buena parte del empleo femenino se concentra aquí, la pregunta clave no es solo “estar” en el sector, sino qué sub-rama y qué formalidad. Las diferencias en calidad del puesto y protecciones siguen explicando asimetrías de retorno dentro del mismo terciario.

(iii) Umbrales salariales y beneficios

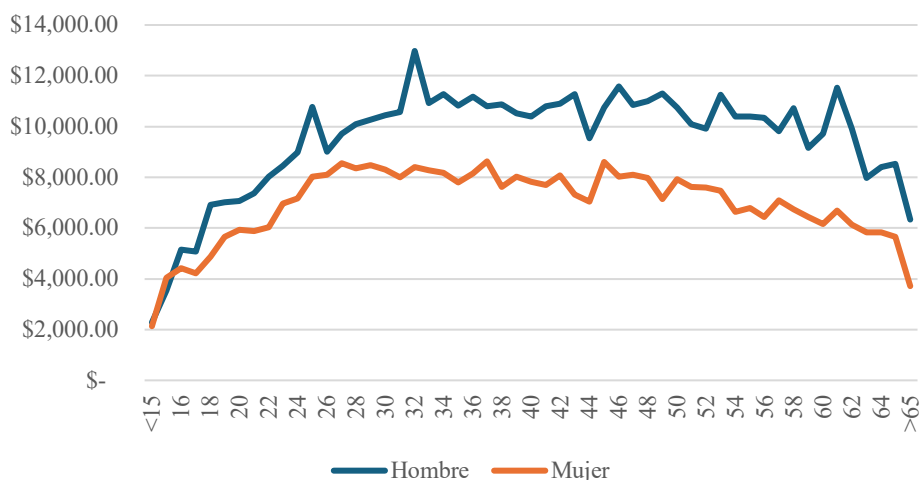
Dummies de umbral ( $\geq 2$  y  $\geq 3$  salarios mínimos). Ambas variables muestran coeficientes positivos muy altos (por construcción, son marcadores de tramos altos del ingreso). Interesa el diferencial por sexo: en el umbral de  $\geq 3$  SM los coeficientes son mayores para mujeres en los tres años, lo que indica que cuando ellas alcanzan esa franja, el “salto” en log-salario es al menos tan grande como en los hombres —pero mucho menos. En suma: no es que las mujeres no reciban prima en la cúpula; es que llegan menos y más tarde.

Servicios médicos. La cobertura está asociada a premios salariales relevantes para ambos, más altos en mujeres; coherente con la prima de formalidad y con el valor de los beneficios no salariales (que suelen correlacionar con mejores escalas de remuneración). La asimetría sugiere que, a igual cobertura, el “paquete” laboral femenino incorpora aumentos en  $\ln(w)$  equivalentes o mayores, aunque la probabilidad de acceso siga siendo menor para ellas en algunos dominios.

(iv) Ciclo de Vida

El perfil por grupos etarios ofrece la curva típica de ingresos-experiencia: aumentos en tramos 14–24, 24–34, 34–44 (pico), y moderación posterior. El máximo salarial se ubica entre 34–44 años para ambos sexos en los tres cortes, con premios significativos; desde allí, los retornos se reducen. Esta trayectoria es compatible con el capital humano acumulado y con transiciones a posiciones de responsabilidad, pero también con rupturas (maternidad/cuidado) que, para mujeres, pueden afectar la continuidad y la velocidad de ascenso.

Gráfico 4.1 Media salarial etaria 2023



Fuente: Creación propia. Media salarial etaria 2023: conformación a través de método de winsorización p1, p99 con selección de salario > 0 con datos de la ENOE 2023 INEGI.

(v) Educación

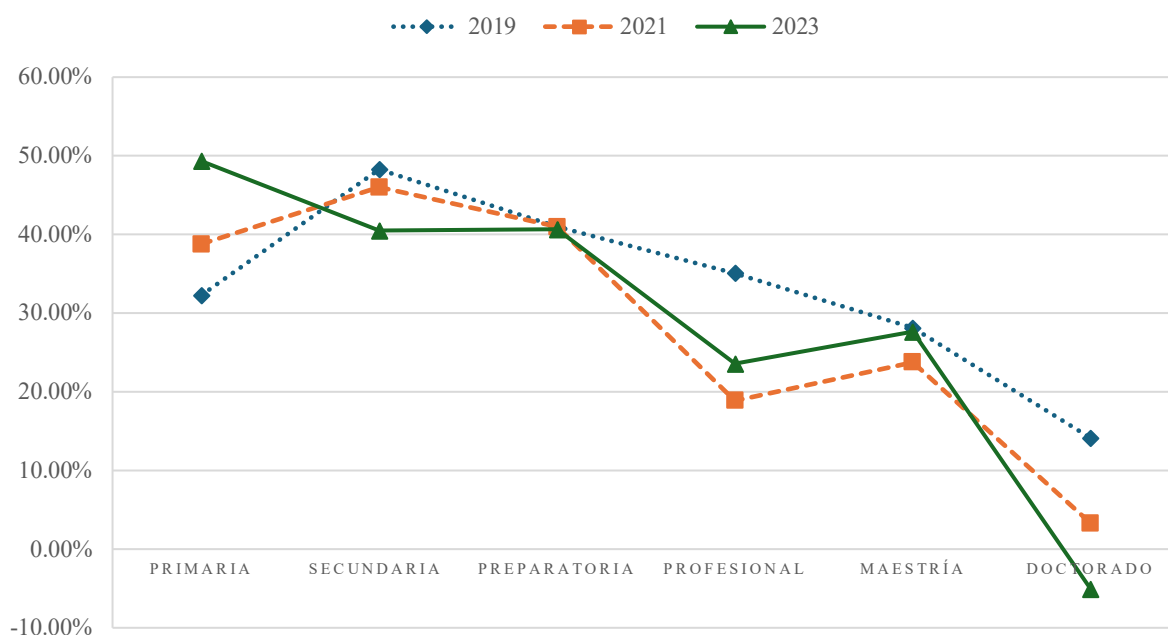
Sin estudios / primaria / secundaria: descuentos pronunciados en ambos sexos, más fuertes en mujeres; la penalización por baja escolaridad amplifica el diferencial de género.

Preparatoria: efectos modestos; en mujeres persiste un castigo pequeño, señalando que el umbral real de “protección” salarial no se consolida en media superior.

Profesional (licenciatura): premios robustos y significativos en ambos; en mujeres el signo pasa de ligeramente negativo en 2019 a claramente positivo en 2021–2023, sugiriendo mejoras en la valorización de credenciales femeninas post-choque.

Posgrado: primas altas y consistentes para ambos; en 2021–2023 el retorno femenino es especialmente elevado, lo que coincide con la evidencia de que una vez en la cúpula educativa, la recompensa existe. La cuestión de fondo, otra vez, es acceso y frecuencia: cuántas mujeres llegan y en qué ramas ocupacionales.

Grafico 4.2 Relación brecha salarial y nivel de estudios  
2019, 2021 y 2023



Fuente: Creación propia. Relación de la brecha salarial con respecto al nivel de estudios, con datos de la ENOE INEGI.

#### (vi) Alfabetización

La no alfabetización implica descuentos grandes y sistemáticos de hecho, entre los más fuertes del modelo, ligeramente más altos para mujeres en 2023. Esto confirma que, en el estrato de menor calificación, se concentran penalizaciones múltiples (sector, formalidad, jornada, protección social), que operan como piso pegajoso.

#### (vii) Estado Civil

Estar casado/en unión, se asocia a premios significativos para los hombres de forma consistente en los tres años mientras que en las mujeres el efecto es nulo o negativo. La divergencia apunta a normas y arreglos del hogar que feminizan el trabajo de cuidado y penalización la trayectoria remunerado femenino, incluso a igual estado conyugal. Este es uno de los canales que reaparecerá con el peso en la descomposición.

### 4.2 Oaxaca Blinder descomposición triple

Tabla 4.4. Descomposición triple O-B, características, coeficientes e interacción en la explicación salarial

Variable	2019			2021			2023		
	Endowments	Coefficients	Interaction	Endowments	Coefficients	Interaction	Endowments	Coefficients	Interaction
<i>Rural</i>	-0.003*** (0.000)	-0.004** (0.002)	-0.001** (0.000)	-0.004*** (0.000)	0.014*** (0.001)	0.002*** (0.000)	-0.003*** (0.000)	0.008*** (0.001)	0.001*** (0.000)
<i>No sabe leer/escribir</i>	-0.000*** (0.000)	-0.000 (0.001)	-0.000 (0.000)	-0.000*** (0.000)	0.001*** (0.000)	0.000*** (0.000)	-0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)
<i>Estado civil</i>	-0.000 (0.001)	0.056*** (0.002)	0.017*** (0.001)	-0.002*** (0.000)	0.047*** (0.002)	0.014*** (0.001)	-0.002*** (0.000)	0.072*** (0.002)	0.019*** (0.001)
<i>Subocupado</i>	-0.004*** (0.000)	0.001 (0.001)	0.000 (0.000)	-0.001*** (0.000)	-0.005*** (0.001)	-0.000*** (0.000)	-0.000 (0.000)	0.005*** (0.001)	0.000 (0.000)
<i>Zona A</i>	-0.001*** (0.000)	-0.004*** (0.001)	0.000*** (0.000)	0.002*** (0.000)	-0.005*** (0.000)	-0.000*** (0.000)	0 (0.000)	-0.006*** (0.000)	-0.000 (0.000)
<i>Act. Primarias</i>	0.014*** (0.004)	-0.004*** (0.001)	-0.018*** (0.005)	0.003 0.002	-0.005*** (0.001)	-0.020*** (0.003)	0.016*** (0.001)	-0.006*** (0.001)	-0.021*** (0.002)
<i>Act. Secundarias</i>	0.014* (0.008)	0.053*** (0.013)	0.041*** (0.010)	-0.017*** (0.004)	0.055*** (0.006)	0.047*** (0.005)	0.021*** (0.003)	0.020*** (0.005)	0.018*** (0.005)
<i>Act. Terciaras</i>	-0.070*** (0.011)	-0.017 (0.050)	0.005 (0.014)	0.008 (0.005)	0.124*** (0.024)	-0.037*** (0.007)	-0.045*** (0.004)	-0.021 (0.024)	0.006 (0.007)
<i>Dos salarios</i>	-0.010*** (0.001)	-0.152*** (0.002)	0.011*** (0.001)	0.009*** (0.001)	-0.128*** (0.002)	-0.007*** (0.000)	0.032*** (0.001)	-0.097*** (0.001)	-0.016*** (0.000)
<i>Mas de tres salarios</i>	0.097*** (0.002)	-0.038*** (0.001)	-0.023*** (0.001)	0.076*** (0.001)	-0.017*** (0.001)	-0.010*** (0.000)	0.061*** (0.001)	-0.009*** (0.000)	-0.007*** (0.000)
<i>Servicios médicos</i>	0.003*** (0.001)	-0.166*** (0.002)	-0.003*** (0.001)	-0.003*** (0.001)	-0.100*** (0.002)	0.002*** (0.000)	0.001** (0.001)	-0.084*** (0.002)	-0.001** (0.000)

Fuente: Creación propia con datos de ENOE 2019, 2021 y 2023 INEGI. Nivel de significancia \* p < 0.05; \*\* p < 0.01; \*\*\*p < 0.001 <sup>26</sup>.

<sup>26</sup> La descomposición Oaxaca-Blinder en su versión triple separa la brecha salarial media en tres componentes

1. Dotaciones (Endowments “E”): cuanto de la brecha se debe a que hombres y mujeres no compartan la misma distribución de características observables (educación, sector, estado civil, etc.);
2. Retornos (Coefficients “C”): cuanto proviene de que el mercado retribuye de manera distinta esas mismas características según el sexo;
3. Interacción (Interaction “I”): el termino de cruce que capta que la brecha no es lineal y depende conjuntamente de diferencias en dotaciones y en coeficientes

En los tres años analizados, la parte de Retornos (C)<sup>27</sup> aporta la mayor fracción positiva a la brecha en varios bloques clave como el estado civil, sectores y umbrales de ingreso), lo que confirma que, a igual característica, el “precio” que paga el mercado favorece a los varones. El componente de interacción (I) es estadísticamente distinto a cero en múltiples variables, como lo es en estado civil, actividades secundarias y umbrales altos, esto indica que la brecha se amplifica cuando coinciden (i) una distribución de dotaciones más favorable a los hombres y (ii) retornos que también les favorecen. En paralelo, algunos E negativos (como la educación, en ciertos niveles o en actividades terciarias,) muestran que la composición por si sola tuviese a reducir la brecha; sin embargo, en C e I revierten ese efecto.

Descripción de variables:

#### (a) Ruralidad

La ruralidad aparece con E negativo y significativo en 2019 y 2021: la distribución de residencia entre sexos no empuja per se la brecha al alza; en cambio, los retornos cambian de signo entre años y el término de interacción es pequeño pero significativo, sugiriendo que cuando la composición se vuelve relativamente menos desfavorable para las mujeres, persisten diferencias de precio que sostienen el diferencial. En 2021 y 2023, el componente C muestra valores positivos y significativos, coherentes con premios relativos masculinos en entornos no rurales o con penalizaciones femeninas más marcadas en zonas rurales.

#### (b) Alfabetización

“No sabe leer/escribir” aporta E muy pequeño y negativo (significativo a niveles altos), mientras que C e I son prácticamente nulos o marginales. La implicación es directa: el analfabetismo diferencia poco la composición por sexo y no introduce grandes asimetrías de precio adicionales en el promedio; su papel está sobre todo en niveles (castigos), más que en la brecha agregada.

---

<sup>27</sup> En notación compacta, si  $\bar{X}_h, \bar{X}_m$  denotan promedios de covariables y  $\hat{\beta}_h, \hat{\beta}_m$  los vectores de retornos estimados por sexo (con  $\hat{\beta}^*$  como estructura de referencia), la diferencia en log-salario  $\Delta \ln \bar{w}_h - \ln \bar{w}_m$  puede expresarse como:

$$\Delta = (\bar{X}_h - \bar{X}_m) \hat{\beta}^* + \bar{X}_m (\hat{\beta}_h - \hat{\beta}_m) + (\bar{X}_h - \bar{X}_m) (\hat{\beta}_h - \hat{\beta}_m)$$

que identificamos como E + C + I, respectivamente. Esta partición permite asilar mecanismos: composición (E), discriminación o retornos asimétricos (C) e interacciones (I).

### (c) Estado civil

Es uno de los motores centrales de la brecha. La parte de retornos (C) es positiva y elevada en todos los años ( $\approx 0.056$  en 2019;  $\approx 0.047$  en 2021;  $\approx 0.072$  en 2023, todos a \*\*\*), y la interacción (I) también positiva y significativa ( $\approx 0.017$ ; 0.014; 0.019). En contraste, las dotaciones (E) son pequeñas o incluso negativas (y significativas en 2021–2023). Traducción económica: aunque la composición de estados conyugales no favorecería por sí sola un aumento de la brecha, el premio conyugal para hombres y su cruce con la distribución ensancha de forma robusta el diferencial. Esto es consistente con penalizaciones por arreglos del hogar y cuidado que no se retribuyen en el caso femenino.

### (d) Subocupación

Patrón no monótono entre años. En 2019, E es negativo y pequeño (\*\*\*), indicando que la mayor presencia femenina en subocupación tendería a reducir la brecha (porque ellas ya están más castigadas en nivel); en 2021 reaparece un C negativo y significativo ( $\approx -0.005$ ), y en 2023 C cambia de signo a positivo y significativo ( $\approx 0.005$ ), con E casi nulo. La lectura es que la “pérdida” por subocupación recae relativamente más en mujeres, y cuando esa asimetría se intensifica, parte de la brecha medida se reduce vía componente C negativo; en 2023, la reconfiguración del mercado hace que el castigo relativo se redistribuya y el efecto vuelva a sumar a la brecha.

### (e) Zona A (ámbitos de alta remuneración relativa)

En 2019, E es negativo, señalando que una mayor presencia masculina en zonas desarrolladas no explica un aumento de la brecha; C también negativo, lo que sugiere premios insuficientes para ambos sexos con un sesgo que no agranda la brecha. A partir de 2021, E pasa a positivo (\*\*\*), de modo que la composición ahora sí empuja la brecha, mientras C sigue siendo negativo y significativo; el cruce (I) es muy pequeño.

### (f) Sectores de actividad

Primarias. En 2019, E positivo, C negativo, I negativo (\*\*\*): la composición (más hombres en el primario) sumaría brecha, pero los retornos e interacción la reducen. Hacia 2023, E se mantiene positivo y significativo, C continúa negativo y I negativo. Conclusión: el sector no es

homogéneo; dónde se ubican hombres y mujeres dentro del primario y cómo se pagan ocupaciones específicas importa más que el rótulo sectorial.

Secundarias (industria). Es el bloque más nítido en favor de la brecha: C es positivo y alto en los tres años ( $\approx 0.053$ ;  $0.055$ ;  $0.020$ , todos) y I también positivo ( $\approx 0.041$ ;  $0.047$ ;  $0.018$ ,) en 2019 el E es positivo y marginal, en 2021 E se vuelve negativo, y en 2023 E positivo (\*). La conclusión es robusta: la industria remunera con primas y los hombres capturan más de esa renta; cuando además se concentran en puestos mejor pagados dentro del sector, la interacción amplifica la brecha.

Terciarias (servicios). Presentan E negativo y significativo en 2019 ( $-0.070$  \*\*\*) y 2023 ( $-0.045$  \*\*\*), es decir, la composición tendería a cerrar la brecha. Sin embargo, en 2021 aparece un C positivo y grande ( $\approx 0.124$  \*\*\*), con I negativo y significativo; la señal de 2021 es consistente con la recomposición pospandemia: a igual inserción en servicios, el precio favoreció a los varones ese año. En 2019 y 2023, al dominar E negativo, el terciario modera la brecha por composición.

#### (g) Umbrales salariales ( $\geq 2$ SM y $\geq 3$ SM)

En los tres cortes, E es negativo para 2 SM ( $-0.010$ ;  $+ 0.009$ ;  $+ 0.032$  con signos que cambian de año a año y alta significancia), y positivo para  $\geq 3$  SM ( $\approx 0.097$ ;  $0.076$ ;  $0.061$ , \*\*\*), indicando que la presencia masculina en la cúpula contribuye a ensanchar la brecha por composición. Por su parte, C es negativo y grande en ambos umbrales y años (pj.,  $-0.152$ ;  $-0.128$ ;  $-0.097$  en “dos salarios”, todos \*\*\*, y  $-0.038$ ;  $-0.017$ ;  $-0.009$  en “ $\geq 3$  SM”, \*\*\*) lo que refleja que, si hombres y mujeres tuvieran idéntica distribución en esos umbrales, la brecha caería por el lado de los retornos; dicho de otro modo, ellas no reciben premios superiores a los de ellos en la cúpula, y C negativo compensa parcialmente lo que E añade. El I alterna en signo, pero suele ser pequeño; sugiere que la combinación de “quién llega” y “cómo se paga” no es lineal y varía por ciclo.

#### (h) Servicios médicos (proxy de formalidad/beneficios)

En 2019, E positivo y pequeño ( $\approx 0.01$ ), con C muy negativo ( $\approx -0.166$ ) y I negativo; en 2021, E negativo ( $\approx -0.01$ ), C continúa muy negativo ( $\approx -0.100$ ), e I ahora positivo; en 2023, E positivo (\*\*), C muy negativo ( $\approx -0.084^*$ ) e I ligeramente negativo. El patrón es inequívoco: la

formalidad/beneficios están fuertemente correlacionados con mejores salarios, pero el componente de retornos no engorda la brecha; más bien, si mujeres y hombres tuvieran igual acceso, C tendería a cerrarla. El problema central no es el precio de la formalidad, sino el acceso/composición (quién la obtiene, cuándo y en qué ocupaciones).

### 4.3 Oaxaca Blinder con corrección por selección Heckman

En contraste de resultados cuando se incorpora la corrección por selección a la ecuación salarial, esto es el esquema de Heckman-Oaxaca<sup>28</sup>, del cual salen dos aportaciones analíticas: depura la comparación de salarios medios entre sexos del sesgo inherente a observar remuneraciones solo entre quienes participan y tienen ingreso reportado; por el otro, permite mantener la descomposición de la brecha en termino de dotaciones y retornos, ahora sobre una ecuación de salarios corregida por el proceso de participación.

Tabla 4.5 Modelo Oaxaca-Blinder con Corrección de Sesgo de Heckman

Año	Grupo	N	R <sup>2</sup>	Adj. R <sup>2</sup>	Wald chi <sup>2</sup>	Prob > chi <sup>2</sup>	Root MSE
2019	Hombres	159,649	0.5176	0.5175	61,317.24***	0	0.4982
	Mujeres	128,747	-	-	-	-	-
2021	Hombres	264,591	0.5021	0.502	97,520.51***	0	0.53349
	Mujeres	209,223	-	-	-	-	-
2023	Hombres	288,979	0.485	0.485	102,122.35***	0	0.51571
	Mujeres	236,937	-	-	-	-	-

Fuente: Elaboración propia con datos de ENOE 2019, 2021 y 2023.

En la práctica, el primer bloque de resultados (estadísticos de ajuste, potencial global del modelo y dispersión residual) muestra que la ecuación para hombres captura de forma estable entre 48.5% y 51.8% de variabilidad del log-salario en 2019-2023, con una ligera tendencia descendente del R<sup>2</sup> ajustado en el tiempo, indicio de que crecen los factores no observables o

<sup>28</sup> El término  $\lambda$  estimado en el modelo de Heckman corresponde a la razón inversa de Mills y resume el sesgo asociado a observar salarios únicamente para quienes participan en el mercado laboral remunerado. Un coeficiente significativo de  $\lambda$  indica que la selección no es aleatoria y que la ecuación salarial estimada sin corrección estaría sesgada. En nuestros resultados, la significancia de este parámetro confirma la presencia de selección muestral, por lo que la comparación de salarios medios entre hombres y mujeres basada en la ecuación corregida constituye una aproximación más adecuada a la brecha salarial neta de este sesgo.

que se ha vuelto más heterogénea la estructura de remuneraciones en la etapa postpandemia. A la vez, el *Wald chi*<sup>2</sup> el indicador mantiene niveles muy elevados y con p-value  $\approx 0$  en todos los cortes de ~61 mil en 2019 a ~102 mil en 2023), lo que confirma significancia conjunta de los regresores y, por ende, relevancia estadística de bloque explicativo: el Root MSE aumenta ligeramente, consistente con mayor dispersión salarial en 2021-2023. En términos sustantivos: el modelo sigue explicando mucho y bien, pero cada vez hace más “ruido” con contenido económico que no se agota con covariables estándar y justifica, por diseño, la corrección de selección y a la descomposición posterior.

Al inspeccionar los coeficientes por sexo y año sobre  $\ln(w)$ , aparecen regularidades robustas. La ruralidad penaliza el salario en ambos grupos a lo largo del período; la magnitud es clara y estable en hombres y muestra castigos relevantes mujeres, especialmente tras 2021, esto sugiere que el mismo retorno territorial se traduce en retornos inferiores para ellas, probablemente por la combinación de menos formalidad, menor densidad de ocupaciones con primas y mayores fricciones de movilidad. El analfabetismo exhibe un descuento muy severo y creciente en el tiempo —uno de los más intensos del modelo—, con efecto particularmente marcado en mujeres al final del período: ello coloca a las trabajadoras con baja calificación en el núcleo del piso pegajoso, donde se acumulan precariedad, baja productividad y menor cobertura de derechos. La subocupación también conlleva pérdidas significativas para ambos sexos (del orden de 13–23% según año y grupo); el patrón interanual revela que la penalización se intensifica en los momentos de recomposición del mercado y golpea con fuerza a las mujeres cuando su ruta ocupacional es más inestable, reforzando la idea de participación condicionada y de menor poder de negociación en la asignación de jornadas. En contraste, residir en “zona A” —proxy de ámbitos urbano-metropolitanos o de mayor dinamismo— se asocia a primas salariales altas y consistentes para hombres y mujeres en los tres cortes, con señales incluso ligeramente mayores para mujeres en algunos años: la clave, como se discutió en el capítulo, no es tanto si hay prima, sino quién accede a los puestos que la pagan.

Tabla 4.6 Regresión O-B Heckman<sup>29</sup>

Variable	2019		2021		2023	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
<i>Rural</i>	-0.056*** (0.0027)	-0.003 (0.0054)	-0.038*** (0.0023)	-0.063*** (0.0038)	-0.033*** (0.0021)	-0.040*** (0.0037)
<i>No sabe leer/escribir</i>	-0.150*** (0.0138)	-0.112** (0.0328)	-0.072*** (0.0116)	-0.218*** (0.0249)	-0.239*** (0.0121)	-0.243*** (0.0224)
<i>Subocupado</i>	-0.162*** (0.0044)	-0.156*** (0.0097)	-0.231*** (0.0031)	-0.128*** (0.0054)	-0.199*** (0.0035)	-0.229*** (0.0064)
<i>Zona A</i>	0.291*** (0.0046)	0.295*** (0.0092)	0.206*** (0.0039)	0.241*** (0.0073)	0.277*** (0.0037)	0.324*** (0.0065)
<i>Act. Primarias</i>	-0.117*** (0.0252)	0.002 (0.0538)	-0.263*** (0.0180)	-0.0001 (0.0390)	-0.114*** (0.0184)	0.246*** (0.0350)
<i>Act. Secundarias</i>	0.385*** (0.0250)	-0.069 (0.0518)	0.203*** (0.0178)	-0.011 (0.0371)	0.252*** (0.0183)	0.131*** (0.0338)
<i>Act. Terciarias</i>	0.309*** (0.0250)	0.241*** (0.0516)	0.129*** (0.0178)	0.071* (0.0369)	0.177*** (0.0183)	0.206*** (0.0336)
<i>Dos salarios</i>	0.034*** (0.0028)	0.415*** (0.0056)	0.195*** (0.0023)	0.432*** (0.0039)	0.319*** (0.0021)	0.502*** (0.0038)
<i>Más de 3 salarios</i>	0.959*** (0.0038)	1.252*** (0.0097)	1.169*** (0.0035)	1.278*** (0.0077)	1.314*** (0.0039)	1.450*** (0.0095)
<i>Sector salud</i>	0.110*** (0.0028)	0.490*** (0.0058)	0.191*** (0.0024)	0.383*** (0.0042)	0.182*** (0.0022)	0.356*** (0.0041)
<i>Menor a 14 años</i>	-0.826*** (0.0156)	-0.506*** (0.0363)	-0.588*** (0.0141)	-0.365*** (0.0220)	-0.804*** (0.0121)	-0.491*** (0.0238)
<i>14 - 24 años</i>	0.059*** (0.0072)	0.277*** (0.0144)	0.044*** (0.0054)	0.177*** (0.0107)	-0.061*** (0.0048)	0.202*** (0.0096)
<i>24 - 34 años</i>	0.294*** (0.0070)	0.455*** (0.0141)	0.227*** (0.0052)	0.364*** (0.0102)	0.145*** (0.0047)	0.326*** (0.0093)
<i>34 - 44 años</i>	0.286*** (0.0069)	0.425*** (0.0140)	0.269*** (0.0051)	0.320*** (0.0100)	0.170*** (0.0047)	0.337*** (0.0091)
<i>44 - 54 años</i>	0.290*** (0.0069)	0.409*** (0.0136)	0.254*** (0.0052)	0.350*** (0.0099)	0.192*** (0.0047)	0.327*** (0.0090)
<i>54 - 64 años</i>	0.190*** (0.0073)	-0.096*** (0.0139)	0.136*** (0.0054)	0.249*** (0.0102)	0.122*** (0.0049)	0.211*** (0.0092)

Fuente: Creación propia con datos ENOE 2019, 2021 y 2023. INEGI. Nivel de significancia \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

<sup>29</sup> De manera econométrica, este bloque aporta parámetros de selección. En la práctica, el modelo reporta el  $\sigma$  (desviación estándar del término de error en la ecuación salarial), el  $\rho$  (correlación entre los errores de selección y de salarios),  $\lambda$  o *inversa de Mills* (igual a  $\sigma\rho$ ), así como el test de independencia entre ambas ecuaciones. Cuando  $\rho \neq 0$  y  $\lambda$  resulta significativo, la selección no es aleatoria: quienes observamos con salario tienen características no observadas correlacionadas con su salario potencial.

Las diferencias por sector confirman la segmentación de retornos. En actividades primarias, los hombres enfrentan descuentos (negativos significativos) en 2019–2023, mientras que las mujeres muestran en 2023 una prima positiva; lejos de implicar una “ventaja sectorial” femenina general, esta señal suele recoger selección hacia nichos específicos dentro del primario (puestos organizados o tareas no manuales con formalidad relativa) y no un premio universal. La industria (secundarias) otorga primas claras en todos los años para hombres, y su señal para mujeres pasa de no significativa/atenuada en 2019–2021 a positiva y significativa en 2023: esto sugiere que, aun cuando ambos se benefician de la puerta industrial, ellos capturan históricamente más renta del sector —consistente con posiciones técnicas y jerárquicas—, y ellas comienzan a capitalizar más en la fase de normalización, aunque partiendo de menor presencia relativa. El terciario mantiene premios positivos en ambos sexos (típicamente menores en 2021 para las mujeres, año de recomposición), lo que obliga a mirar sub-ramas y calidad del puesto: no toda “actividad de servicios” paga igual ni retribuye del mismo modo a hombres y mujeres con iguales credenciales. En el bloque de umbrales salariales, las dummies de  $\geq 2$  SM y  $\geq 3$  SM muestran, como es esperable, coeficientes muy altos y significativos; destaca que en  $\geq 3$  SM el salto estimado para mujeres es tan grande o mayor que el de hombres en los tres años: cuando ellas llegan a la cúpula, la prima existe.

Tabla 4.6.2 Regresión O-B Heckman

Variable	2019		2021		2023	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
<i>Soltera, sin hijos</i>	-	0.827*** (0.0142)	-	1.008*** (0.0125)	-	0.999*** (0.0119)
<i>Un hijo</i>	-	-0.068*** (0.0147)	-	0.124*** (0.0130)	-	0.116*** (0.0123)
<i>2 hijos</i>	-	-0.414*** (0.0137)	-	-0.196*** (0.0123)	-	-0.231*** (0.0117)
<i>3 hijos</i>	-	-0.604*** (0.0141)	-	-0.237*** (0.0127)	-	-0.177*** (0.0121)
<i>4 hijos o mas</i>	-	-0.375*** (0.0173)	-	-0.197*** (0.0148)	-	-0.131*** (0.0144)
<i>Constante</i>	7.964*** (0.0269)	7.530*** (0.0547)	8.258*** (0.0196)	7.716*** (0.0393)	8.447*** (0.0198)	7.759*** (0.0359)

Fuente: Creación propia con datos ENOE 2019, 2021 y 2023. INEGI. Nivel de significancia \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\* $p < 0.001$ .

El problema no es el precio del “club de altos ingresos”, sino la subrepresentación femenina en esa franja (lo ya documentado con el top-10%), que mantiene pequeña la contribución de esa prima al promedio agregado femenino. En la misma línea, la variable de servicios médicos — proxy de formalidad/beneficios— se asocia a premios elevados para ambos, generalmente mayores en mujeres: a igual cobertura, el paquete laboral femenino sí sube el log-salario; el cuello de botella está en el acceso diferencial a esa cobertura.

La curva ciclo-vida (grupos de edad) conserva la forma minceriana: ascenso de 14–24, salto en 24–34 y pico en 34–44, con moderación posterior; el patrón se replica en ambos sexos en 2019–2023, aunque con diferencias en magnitud que sugieren trayectorias de acumulación y permanencias dispares (interrupciones por cuidados, retornos decrecientes por techos organizacionales). En el bloque conyugal y fecundidad aparece uno de los mecanismos más contundentes de divergencia: al interior del grupo femenino, ser soltera sin hijos se asocia a un mejor posicionamiento salarial, mientras que la presencia de hijos genera penalizaciones crecientes a partir del segundo/tercer hijo, consistentes en los tres años. El patrón no se reproduce con la misma fuerza (ni dirección) en hombres, y ofrece una lectura clara: el estado civil y la maternidad operan como mecanismos institucionales de freno a la trayectoria de ingresos de las mujeres, incluso controlando por educación, sector y territorio. Así, la evidencia por coeficientes refuerza la narrativa combinada de piso pegajoso (penalizaciones en baja calificación, ruralidad, subocupación y ausencia de beneficios) y techo de cristal (primas en industria, zonas dinámicas y cúpula salarial a las que menos mujeres acceden), claves para entender por qué la porción no explicada domina en la descomposición.

Este ejercicio de Heckman corrección por selección de muestra confirma que la desigualdad de género: se ve reducida parcialmente al depurar la composición de quien logra estar con salario observado, pero permanece un diferencial amplio que la descomposición atribuye, mayoritariamente, a retornos asimétricos y, en segundo término, a interacciones donde dotaciones y precios se alinean a favor de los varones. Importantes observaciones podemos rescatar, por ejemplo: los mecanismos de piso pegajoso y techo de cristal, debe ser considerado en política pública para tomar en cuenta que no es “quienes son” (dotaciones), sino como se paga (retornos), y quienes acceden a puestos con primas, entonces las medidas deben combinar

igualdad de oportunidades con transparencia salarial, así como políticas de acceso a los segmentos que capturan las rentas más elevadas del mercado.

#### 4.4 Heckman Oaxaca y descomposición triple

Esta última sección, bajo el mismo hilo argumental, la corrección de selección tipo Heckman y la descomposición triple Oaxaca-Blinder. Visto desde el punto metodológico: si los salarios observados dependen de quien logra participar y reportar ingreso, compara medias entre hombres y mujeres sin modelar ese proceso puede pesar tanto niveles como brechas.

Las estimaciones confirman que la selección no es aleatoria, ya que el parámetro  $\lambda$  resulta significativo en las especificaciones, de modo que la ecuación salarial corregida es el soporte adecuado para descomponer la brecha sin confundir composiciones en el acceso con retornos salariales. Además, en este bloque se ve documentado las diferencias notables de retorno por sexo en covariables, por ejemplo, que aun cuando depuramos la selección persisten penalizaciones que el mercado impone con mayor intensidad a las mujeres.

Tabla 4.7. Blinder-Oaxaca Descomposición

	2019	2021	2023
<i>Observaciones</i>	288,396	473,814	525,916
<i>Modelo</i>	Lineal	Lineal	Lineal
<i>Grupo 1: Hombres (sex = 1)</i>	159,649	264,591	288,979
<i>Grupo 2: Mujeres (sex = 2)</i>	128,747	209,223	236,937
<i>Observaciones seleccionadas</i>	51,486	86,311	94,664

Fuente: Creación propia con datos de ENOE 2019, 2021 y 2023 INEGI.

Previo al análisis de variables es conveniente revisar el marco muestral, que da sustento estadístico a las inferencias. El resumen de casos para la descomposición O-B muestra un aumento progresivo del tamaño de la muestra total y, a la vez, del número de observaciones por sexo, en 2019, 2021 y 2023; en paralelo, las “observaciones seleccionadas” (mismas que cumplen con el criterio de selección, y corrección), pasan de cerca de 51 mil en 2019 a casi 95 mil en 2023, señala una mayor cobertura y depuración de datos para la etapa de estimación. El modelo empleado es lineal en los tres cortes, mantenido coherencia de especificación a lo largo

del período. Este cuadro muestra que **la brecha no descansa en un año atípico** ni en tamaños reducidos, evidencia de un andamiaje de datos crecientes amplios y consistentes.

Bajo este soporte, la descomposición triple identifica cuanto de la brecha obedece a dotaciones (E), cuanto a retornos (C) y cuanto al termino de interacción (I). La evidencia es contundente: en la mayor parte de los bloques relevantes, C es el componente que más alimenta la brecha; es decir, a la igual característica observable, el “precio” que paga el mercado favorece mas a los varones. Ello no significa que las dotaciones carezcan de papel; por el contrario, se observan E negativos en variables como educación (en ciertos niveles) o servicios (terciarios) en años específicos, que tendrían a cerrar el diferencial si los precios fuesen simétricos. El problema es que C e I revierten ese efecto: allí donde los hombres concentran mejores dotaciones y además reciben mayores retornos, el termino de interacción potencia la brecha. El resultado es estable entre 2019, y 2023 y no se desdibuja con la recuperación postpandemia.

Tabla 4.8. O-B Triple Descomposición

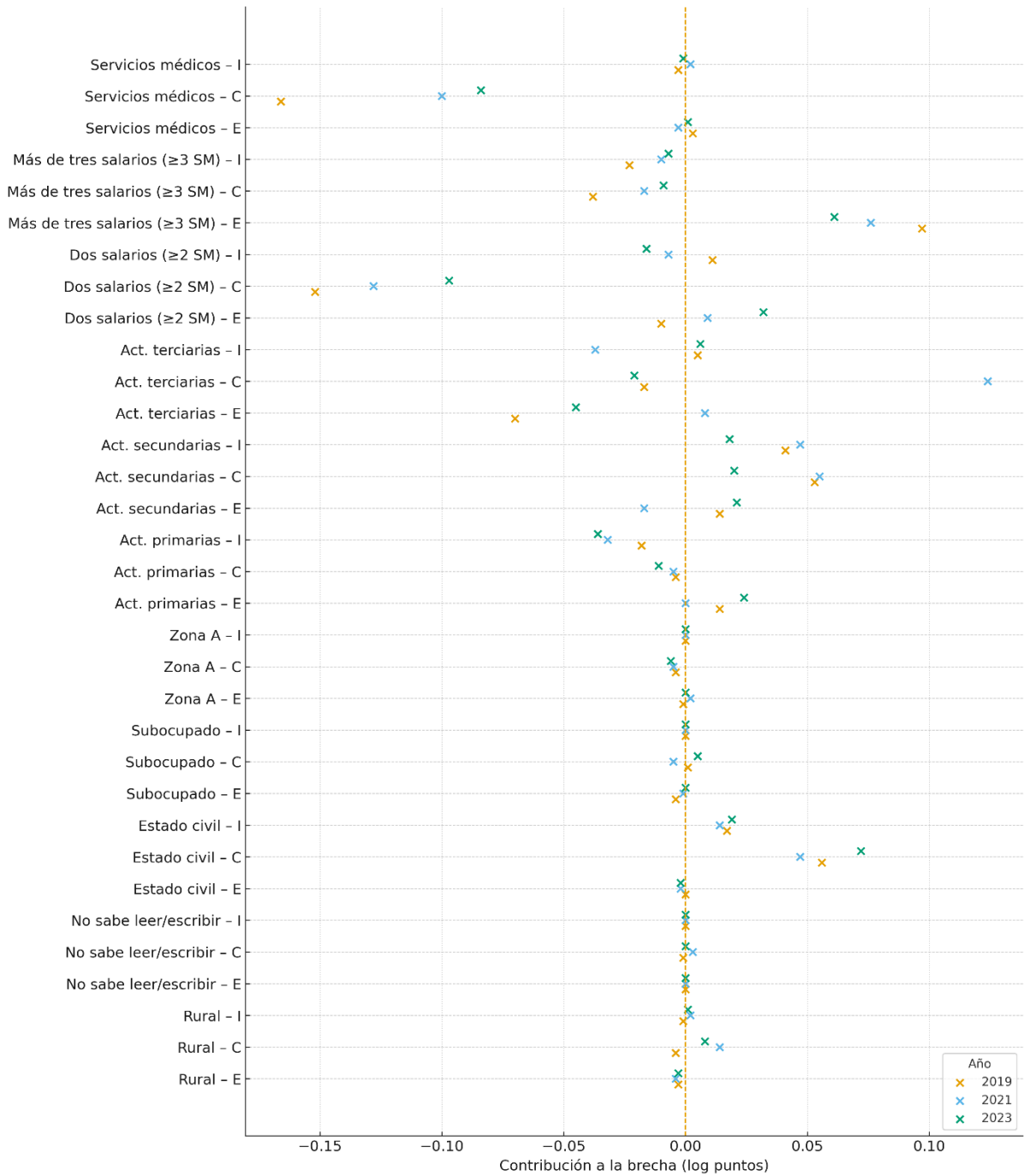
Variable	2019			2021			2023		
	Endowments	Coefficients	Interaction	Endowments	Coefficients	Interaction	Endowments	Coefficients	Interaction
<i>Rural</i>	-0.003*** (0.000)	-0.004** (0.002)	-0.001** (0.000)	-0.004*** (0.000)	0.014*** (0.001)	0.002*** (0.000)	-0.003*** (0.000)	0.008*** (0.001)	0.001*** (0.000)
<i>No sabe leer/escribir</i>	0.000*** (0.000)	-0.001 (0.001)	0.000 (0.000)	-0.000 (0.000)	0.003*** (0.001)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
<i>Subocupado</i>	-0.002*** (0.000)	-0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	-0.014*** (0.001)	-0.000 (0.000)	-0.001*** (0.000)	0.002*** (0.001)	0.000 (0.000)
<i>Zona A</i>	-0.001** (0.000)	-0.000 (0.001)	0.000 (0.000)	0.002*** (0.000)	-0.002*** (0.001)	-0.000 (0.000)	-0.001 (0.000)	-0.004*** (0.001)	0.000 (0.000)
<i>Act. Primarias</i>	0.000 (0.005)	-0.004** (0.002)	-0.011** (0.005)	-0.000 (0.005)	-0.005*** (0.001)	-0.032*** (0.005)	0.024*** (0.003)	-0.011*** (0.001)	-0.036*** (0.004)
<i>Act. Secundarias</i>	-0.008 (0.006)	0.111*** (0.014)	0.053*** (0.007)	-0.002 (0.006)	0.041*** (0.008)	0.034*** (0.006)	0.022*** (0.006)	0.019*** (0.007)	-0.023 (0.008)
<i>Act. Terciarias</i>	-0.050*** (0.011)	0.050 (0.041)	-0.014 (0.012)	-0.020 (0.010)	0.046 (0.032)	-0.016 (0.011)	-0.054*** (0.009)	-0.023 (0.030)	0.008 (0.010)
<i>Dos salarios</i>	-0.021*** (0.001)	-0.170*** (0.003)	0.019*** (0.001)	0.007*** (0.001)	-0.109*** (0.002)	-0.004*** (0.000)	0.039*** (0.001)	-0.073*** (0.002)	-0.014*** (0.001)
<i>Tres salarios o mas</i>	0.100*** (0.002)	-0.029*** (0.001)	-0.023*** (0.001)	0.076*** (0.001)	-0.008*** (0.001)	-0.007*** (0.001)	0.061*** (0.001)	-0.005*** (0.000)	-0.006*** (0.000)
<i>Servicios Medicos</i>	-0.004*** (0.001)	-0.164*** (0.003)	0.003*** (0.001)	-0.014*** (0.001)	-0.085*** (0.002)	0.007*** (0.000)	-0.006*** (0.001)	-0.074*** (0.002)	0.003*** (0.000)
<i>Menor a 14 años</i>	-0.001*** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)	0.001*** (0.000)	-0.002*** (0.000)	0.001*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.000 (0.000)
<i>14 - 24 años</i>	-0.016*** (0.001)	-0.051*** (0.004)	0.013*** (0.001)	-0.011*** (0.001)	-0.031*** (0.003)	0.008*** (0.001)	-0.013*** (0.001)	-0.063*** (0.003)	0.017*** (0.001)
<i>24 - 34 años</i>	0.009*** (0.001)	-0.035*** (0.003)	-0.003*** (0.000)	-0.003*** (0.001)	-0.033*** (0.003)	0.001*** (0.000)	0.003*** (0.001)	-0.041*** (0.002)	-0.002*** (0.000)
<i>34 - 44 años</i>	0.032*** (0.001)	-0.022*** (0.002)	-0.011*** (0.001)	0.011*** (0.001)	-0.010*** (0.002)	-0.002*** (0.000)	0.014*** (0.001)	-0.030*** (0.002)	-0.007*** (0.000)
<i>44 - 54 años</i>	0.002*** (0.001)	-0.022*** (0.003)	-0.001** (0.000)	0.005*** (0.001)	-0.017*** (0.002)	-0.001*** (0.000)	0.003*** (0.001)	-0.024*** (0.002)	-0.001*** (0.000)

Fuente: Creación propia con datos de ENOE 2019, 2021 y 2023 INEGI. Nivel de significancia \* p < 0.05; \*\* p < 0.01; \*\*\*p < 0.001.

Estos bloques ayudan a fijar la idea. Primero, el territorio y las condiciones del puesto, la residencia rural y la subocupación mantienen impactos negativos consistentes en los ingresos, y el peso mayor recae en los coeficientes; eso sugiere que el mercado descuenta con mayor dureza a las mujeres cuando estas condiciones están presentes. El caso de servicios médicos (proxy de formalidad/beneficios) es ilustrativo: porque aun cuando en algunos cortes la composición no agranda e incluso la reduce, la forma en que se remunera la cobertura captada por C no compensa la desventaja agregada, de modo que la desigualdad se sostiene por la vía del “precio” y del acceso diferenciado. Segundo, los umbrales salariales ( $\geq 2$  y  $\geq 3$  SM) dejan un matiz importante: tener tres o más salarios mínimos aparece asociado a una reducción de la brecha por dotaciones (E positivo para varones que explica parte del diferencial), pero acompañado de efectos negativos en retornos para las mujeres; traducido; cuando ellas llegan a la cúpula, el premio existe, pero la frecuencia de acceso y la estructura de pagos no bastan para cerrar la distancia promedio. Tercero, el ciclo de vida; franja 34-44 años aporta de modo significativo a la desigualdad, reflejando beneficios salariales mayores para hombres en ese rango precisamente el momento en que se consolidan posiciones de responsabilidad y se acentúan las penalizaciones por cuidado en trayectorias femeninas. En conjunto, la triple lectura no reduce la brecha a “quien es quien”, sino a como se paga y cuando se permite llegar a los tramos de mayor demanda.

La brecha salarial de género persiste y su corazón se ubica en retornos a los varones. Las covariables que con más fuerza sostienen esta dinámica, ruralidad subocupación, beneficios/formalidad, umbrales de salarios altos, y franja etaria central, remiten a fenómenos que la literatura identifica como piso pegajoso (penalizaciones intensas en la base por baja calificación informalidad y cuidado) y techo de cristal (subrepresentación y premios diferenciales en la cúpula).

Gráfico 4.3. Coefplot Contribuciones E/C/I por variable (2019, 2021, 2023).



Nota: El gráfico muestra las contribuciones relativas de cada variable al componente Explicado (E), Características (C) e Inobservable (I) de la brecha salarial, estimadas mediante la descomposición Oaxaca–Blinder para los años 2019, 2021 y 2023. Los puntos representan la magnitud del aporte de cada factor al diferencial del log-salario entre hombres y mujeres. Fuente: creación propia con datos de la ENOE.

#### 4.5 Modelo espacial SLX (panel municipal): efectos fijos locales y de vecindad.

El bloque espacial se estimó con un modelo SLX en panel balanceado a nivel municipal, lo que permite distinguir entre efectos locales (coeficientes  $\beta$  sobre la brecha “intramunicipal”) y efectos de contexto territorial (coeficientes  $\theta$  sobre los rezagos espaciales de esas mismas variables en la vecindad). Se trabajó con dos especificaciones (efectos fijos “within” y efectos aleatorios “random” y la elección de contraste con test de Hausman. Aun cuando la prueba reporta un estadístico bajo ( $\approx 0.1408$ ), y por lo tanto no rechaza la consistencia de efectos aleatorios, la narrativa principal adopta efectos fijos con prudencia: en este problema es plausible que existan rasgos inobservables y persistentes de cada municipio (trayectoria productiva, instituciones locales, arreglo del mercado de trabajo) correlacionados con las covariables, de modo que fijarlos evita sesgos de estimación.

Tabla 4.9 Estimaciones de las brechas con efectos fijos y aleatorios

Variable	Efectos Fijos	Efectos Aleatorios	Variable	Efectos Fijos	Efectos Aleatorios
<i>esc_gap</i>	0.0482***	0.048***	<i>WX_edad_gap</i>	-0.0015	-0.0015
	-0.0027	-0.0024		-0.0012	-0.0011
<i>edad_gap</i>	0.0031***	0.0023***	<i>WX_formal_gap</i>	0.1233**	0.0841*
	-0.0007	-0.0006		-0.0397	-0.0359
<i>formal_gap</i>	0.5585***	0.5622***	<i>WX_urbano_gap</i>	-0.0016	-0.0406
	-0.0221	-0.02		-0.1034	-0.0947
<i>urbano_gap</i>	0.1879**	0.1584**	<i>WX_rama_industria_gap</i>	-0.0267	-0.0183
	-0.0612	-0.0564		-0.0482	-0.0413
<i>rama_industria_gap</i>	-0.0516	-0.0639**	<i>WX_rama_comercio_gap</i>	-0.042	-0.0462
	-0.0277	-0.0243		-0.039	-0.0339
<i>rama_comercio_gap</i>	0.0492*	0.0434*	<i>Intercepto</i>		0.4019***
	-0.0232	-0.0204			-0.0127
	-0.0075	-0.006			
<i>WX_esc_gap</i>	-0.0046	-0.0042	<i>Hausman</i>	0.1408	No se rechaza H <sub>0</sub> : los efectos aleatorios son consistentes

Modelo SLX en panel municipal balanceado; brecha logarítmica como dependiente; rezagos espaciales construidos con matriz de contigüidad estandarizada por filas. Sig. \*\*\* P<0.001; \*\* P<0.01; \* P<0.05  
Fuente: Elaboración propia con microdatos de la ENOE (2018–2024)

La base se preproceso como panel balanceado, con control de faltantes, y el conjunto de regresores incorpora brechas de escolaridad, edad, formalidad, urbanización y estructura

sectorial (agropecuario, industria y servicios), además de sus rezagos espaciales (WX). Esta arquitectura capta con claridad como pesa el propio municipio y que hereda de su entorno inmediato en a la formación de la brecha salarial por género.

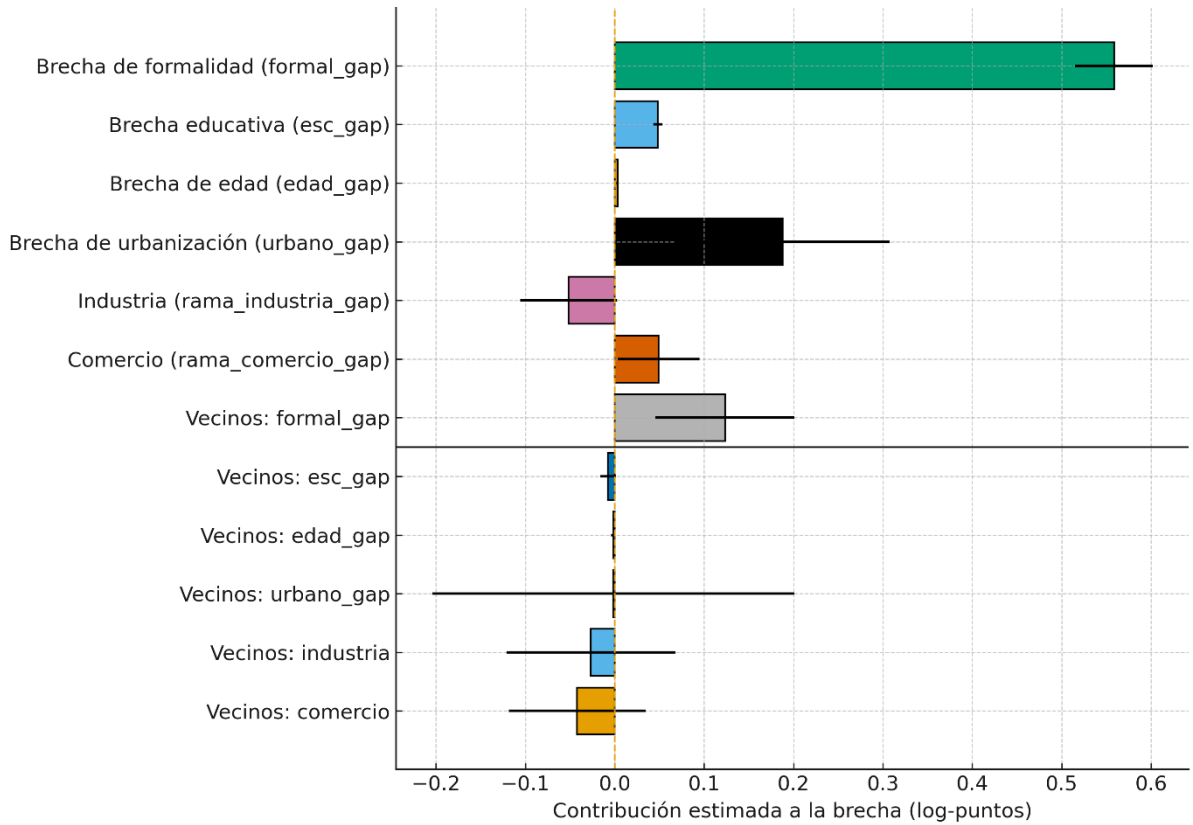
Los efectos locales ( $\beta$ ) trazan una jerarquía nítida de determinantes. La brecha de formalidad surge como el principal motor de la desigualdad salarial: el coeficiente se sitúa en torno a 0.56 tanto en fijos como en aleatorios; en términos sustantivos, un incremento de 10 puntos porcentuales en formalidad masculina relativa eleva la brecha logarítmica del salario en  $\sim 5.6\%$  esto no es un matiz; sintetiza que las puertas de acceso a empleos formales (con seguridad social, estabilidad y bonos) están más abiertas para varones, y que esa asimetría se traduce en brecha agregada incluso tras controlar por otras covariables. La brecha educativa también suma con un coeficiente de  $\approx 0.048$  (cada año adicional de ventaja masculina en escolaridad aumenta la brecha en torno a  $4.8\%$ ), mientras que la diferencia de edad aporta un efecto positivo pero pequeño ( $\approx 0.003$  por año), consistente con un diferencial de ciclo de vida: los municipios donde la edad promedio de los hombres es ligeramente mayor concentran ocupaciones mejor pagadas (experiencia acumulada), pero su contribución es mucho menor que la de formalidad y educación.

La urbanización relativa (mayor presencia masculina en zonas urbanas) aparece con un efecto positivo y significativo, en el orden de 0.16-0.19. La lectura puede ser engañosa, porque el rango efectivo de variación es acotado, pero la señal es coherente: los mercados laborales urbanos pagan primas y ofrecen trayectorias de ascenso que no están distribuidas de manera neutral por sexo; cuando los varones acaparan los puestos urbanos con mejor remuneración, el promedio municipal se abre a su favor. Por estructura productiva, comercio muestra un coeficiente positivo de ( $\approx 0.049$ ), señalando que las configuraciones donde ellos dominan ese segmento, la brecha tiende a ensancharse; en contraste, el bloque industrial arroja un coeficiente negativo y, en general, de significancia débil, en fijos (más claro en aleatorios), lo que sugiere que la industria no necesariamente magnifica la desigualdad municipal : allí donde la manufactura es relativamente formalizada y con reglas más estandarizadas, las diferencias de precio por sexo no se agrandan en promedio, o lo hacen menor que en otros entornos. Este perfil de  $\beta$  es estable entre especificaciones y respalda que los canales de precio centrales (formalidad, educación y urbanización), son los que explican la mayor parte de la variación espacial de la brecha.

Son los efectos de vecindad ( $\theta$ ) el resultado más contundente es el rezago espacial de la formalidad (WX\_formal\_gap), positivo y significativo, en ambas especificaciones ( $\approx 0.12$  en fijos;  $\approx 0.08$  en aleatorios). La interpretación es directa: aun controlando por la estructura del municipio, la desigualdad de formalidad en los municipios vecinos “se derrama” sobre la brecha salarial local. Donde alrededor predomina el acceso formal masculino, el municipio “focal” hereda parte de esas asimetrías, ya sea por mercados de trabajo integrados (conmutación diaria, cadenas de valor) p por mimetismo institucional en la contratación. En cambio, los demás rezagos (educación, edad, urbanización, industria y comercio) no alcanzan significancia de manera sistémica, lo que refuerza que la formalidad es el canal espacial clave en este período: es el “vehículo” que difunde, por proximidad geográfica, la desigualdad de género en salarios. Este hallazgo encaja con los resultados del modelo Oaxaca, donde no basta con la acotación, importa el acceso a posiciones con prima y como el entorno territorial la condiciona.

El ajuste global de los modelos y la significancia conjunta de los regresores resultan consistentes con la validez del enfoque, y el intercepto positivo en fijos, estimados, sugiere un “nivel base” de brecha que luego se amplifica o mitiga por los canales económicos descritos. La nota metodológica de test de Hausman clarifica el porqué de privilegiar fijos en el cuerpo del texto: si los efectos individuales (municipio) se correlacionan con los regresores, los fijos son el estimador consistente, el hecho de que el estadístico no rechace aleatorios señala estabilidad de signos, reporta ambos muestra de robustez, pero no valida la precaución econométrica de controlar por heterogeneidad inobservable con fijos.

Gráfico 4.4. Coefplot de efectos locales ( $\beta$ ) y de vecindad ( $\theta$ ) con IC95%. Modelo SLX de efectos fijos, panel municipal (2019–2023)



Fuente: Creación propia con datos de la ENOE. Modelo SLX en panel municipal. Barras horizontales con IC95%. Los efectos locales de formalidad, educación y urbanización amplían la brecha; entre los rezagos, solo WX\_formal\_gap resulta significativamente positivo, indicando derrames territoriales de la desigualdad de acceso a empleos formales.

En suma, el conjunto de resultados no depende de una sola especificación ni de una sola variable: emerge un patrón estructural donde la brecha de formalidad pesa localmente y por contagio, la adecuación agrega, la urbanización inclina la cancha y el resto actúa en segundo plano o son señales parciales.

El piso aparece donde la formalidad femenina es escasa y, además, las vecindades refuerzan esa escasez; en esos cinturones, las mujeres quedan atrapadas en ocupaciones sin prestaciones ni trayectoria, y el promedio municipal lo refleja. El techo se reconoce en entornos urbanos y terciarios donde, pese a existir primas, ellas acceden menos; cuando lo hacen, el precio existe, pero la frecuencia de acceso no alcanza para cerrar el diferencial agregado. La coincidencia

entre  $\beta_{\text{formal\_gap}}$  grande y  $\theta_{\text{WX\_formal\_gap}}$  significativo deja una enseñanza operativa: las brechas no son meramente “locales”, sino territoriales; se producen en un municipio y se reproducen en su contorno. Para política pública, esto implica que igualar precios sin abrir la puerta formal en los territorios —y en sus redes de vecindad— tendrá impacto limitado.

NOTA técnica: Reporte de FE y RE si Hausman no rechaza

En paneles municipales es razonable suponer que existen rasgos inobservables persistentes (historia productiva, reglas locales de contratación, capacidad institucional) correlacionados con nuestras covariables. Cuando eso ocurre, los efectos fijos (FE) son el estimador consistente porque eliminan, por construcción, esa heterogeneidad no observada. Que la prueba de Hausman no rechace los efectos aleatorios (RE) solo indica que, con la muestra actual, no hay evidencia estadística suficiente para afirmar que RE sea inconsistente; no equivale a demostrar ausencia de correlación con inobservables.

Reportar ambas especificaciones (FE y RE) es una práctica de robustez y transparencia: permite comprobar que los signos y magnitudes clave no dependen de una sola elección econométrica y explicitar al lector qué asume cada estimador. En la narrativa sustantiva privilegiamos FE por prudencia —es la opción más restrictiva respecto a la correlación con inobservables—, y dejamos RE como contraste útil cuando el objetivo es evaluar estabilidad de resultados bajo supuestos menos exigentes.

#### 4.5.1 Ajuste del modelo y validación espacial (2018 – 2024)

El desempeño del SLX a lo largo del período confirma que se trata de una especificación parsimoniosa pero informativa para leer la geografía de la brecha. El  $R^2$  anual fluctúa en una banda 0.167 – 0.268, con un máximo en 219 (0.2683) y caídas visibles en 2020 - 2021. Esta pauta tiene una interpretación económica clara: antes del shock sanitario el sistema municipal presentaba regularidades más estructuradas, de modo que una combinación de brechas locales y de entorno explicaba mayor parte de la variación; con la pandemia, irrumpiendo dinámicas transitorias (cierres, rotaciones, cambios de composición sectorial) que desalinearon los patrones previstos y redujeron la capacidad explicativa de regresores esencialmente

estructurales. Retornos hacia niveles intermedios en 2022 en sugiere que, aunque los mercados locales se estabilizaron, persisten heterogeneidades no observadas que no se agotan con variables estándar. En todo caso, los niveles de  $R^2$  reportados son consistentes con lo esperable en ejercicios municipales, donde una fracción no menor de la varianza reside en normas locales, redes sociales y arreglos institucionales, difícilmente medibles con precisión anual.

Tabla 4.10 Estadísticos de ajuste del modelo SLX por año (2018–2024)

<b>Año</b>	<b>R2</b>	<b>AIC</b>	<b>BIC</b>	<b>Moran_I</b>	<b>p_valor_Moran</b>
2018	0.2088	1226.52	1296.57	-0.0229	0.8185
2019	0.2683	1164.43	1234.16	0.003	0.4346
2020	0.1799	1717.96	1787.72	-0.0403	0.9457
2021	0.1668	1529.68	1599.67	-0.0304	0.8901
2022	0.2137	1084.07	1153.77	-0.0388	0.9394
2023	0.1824	1179.3	1249.29	0.0282	0.109
2024	0.1825	1399.42	1470.12	-0.0047	0.5663

Fuente: Elaboración propia con base en microdatos armonizados de la ENOE, 2018–2024. Cálculos realizados mediante estimaciones anuales del modelo SLX con rezagos espaciales, la ausencia de autocorrelación residual ( $p > 0.1$ ) valida la especificación espacial).

Los criterios de información AIC y BIC acompañan esta lectura: los años mejor ajustados exhiben penalizaciones más bajas, señal de que el modelo no incurre en sobreajuste y mantiene un equilibrio razonable entre simplicidad y poder explicativo. En otras palabras, cuando la coyuntura es menos turbulenta, la misma estructura SLX extra más señal sin necesidad de “inflar” la especificación; cuando el retorno es disruptivo, AIC/BIC, suben por la menor capacidad de variables para “contener” shocks idiosincráticos. Este comportamiento paralelo de ajuste y penalización respalda la adecuación funcional del modelo en todo el período y, para nuestra ventana de resultados, garantiza comparabilidad métrica.

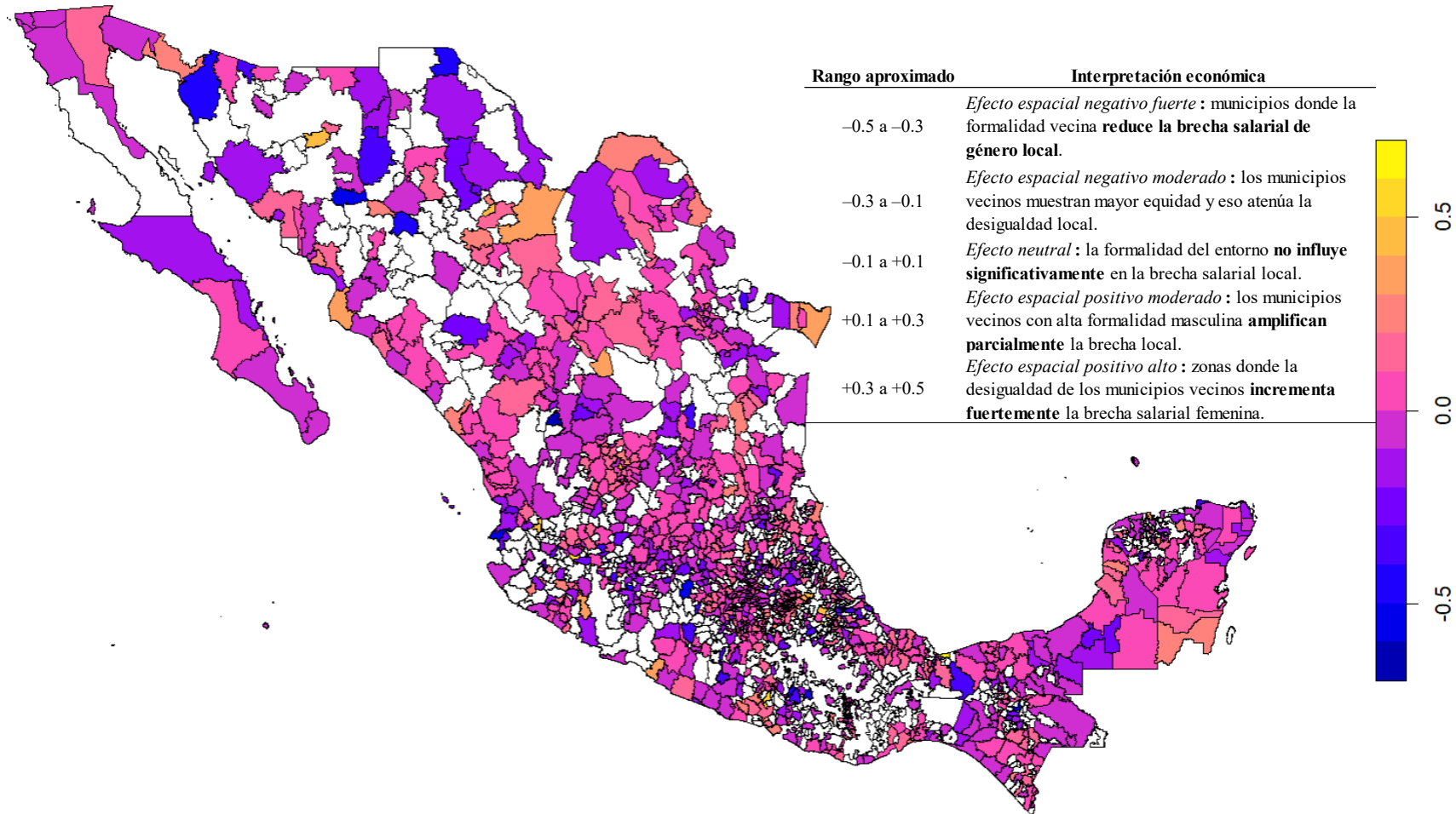
El test de Moran sobre residuos es decisivo para la validez espacial: en todos los años los p-valores superan 0.1, por lo que no se detecta autocorrelación residual. La conclusión es que los rezagos espaciales incorporados por SLX absorben la dependencia geográfica que podría sesgar los coeficientes, evitando patrones espaciales remanentes no explicados. Dicho de otro modo, la estructura de vecindad relevante quedó capturada en el lado derecho de la ecuación y no persiste “geografía en los errores”. Esta propiedad es la que permite interpretar de manera limpia los efectos de contexto estimados y respaldar inferencias territoriales con confianza.

#### 4.5.2 Cartografía de rezago de formalidades y dinámica de clústeres.

El análisis de `WX_formal_gap` (rezago espacial de la brecha de formalidad, ofrece el hallazgo clave del componente territorial: es el único efecto de vecindad que aparece significativo y estable, lo que implica que entornos con alta desigualdad de acceso formal para hombres tienden a amplificar la brecha salarial en el municipio vecino, aun controlando por las propias características locales. La lectura sustantiva es directa: la formalidad opera como el vehículo de transmisión territorial de la desigualdad de género; la geografía de la seguridad social. La estabilidad contractual y las primas asociadas no se queda “dentro” de cada municipio, sino que se difunden por la red de vecindad.

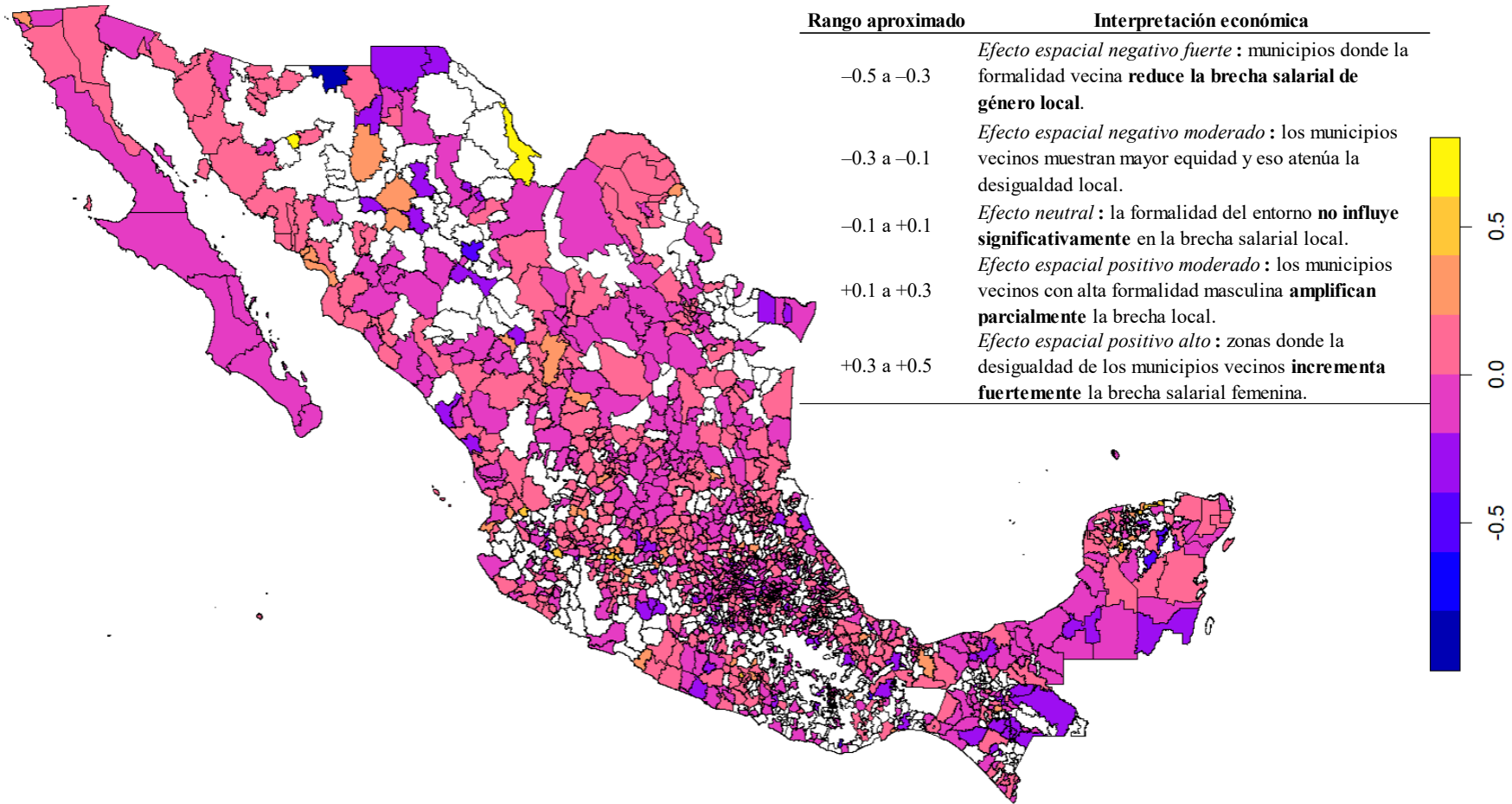
Los mapas anuales 2018 – 2024 confirman esta interpretación. Se observa persistencia espacial: muchos municipios mantienen patrones de rezago similares año tras año, evidencia de rigidez estructural en la organización regional del empleo formal. Tras 2020 emergen clústeres más intensos, (agrupamiento de municipios con `WX_formal_gap` alto) que señalan mecanismos de difusión de la desigualdad, compatibles con mercado laborales integrados (conmutación, cadenas de valor y con prácticas de contratación que se mimetizan a nivel regional. La pandemia, además, dejó un rastro de dispersión y fragmentación temporal del patrón, pero las tendencias estructurales reaparecen conforme la economía local se reordena. Este comportamiento espacial (picos de dispersión de la crisis, vuelta a la pauta de clústeres después) refuerza que el canal de formalidad es el eje territorial de la brecha en el período.

Mapa 1. Rezago espacial de la formalidad (WX\_formal\_gap), 2018.



Fuente: Elaboración propia con base en microdatos armonizados de la ENOE (2018) e información geoespacial del Marco Geoestadístico INEGI 2020. Estimaciones obtenidas mediante modelo SLX con rezago espacial de formalidad (WX\_formal\_gap). Nota: municipios en blanco sin información.

Mapa 1. Rezago espacial de la formalidad (WX\_formal\_gap), 2024.



Fuente: Elaboración propia con base en microdatos armonizados de la ENOE (2024) e información geoespacial del Marco Geoestadístico INEGI 2020. Estimaciones obtenidas mediante modelo SLX con rezago espacial de formalidad (WX\_formal\_gap). Nota: municipios en blanco sin información.

#### 4.5.2.1 Nodos territoriales y varianza de la desigualdad (SLX, 2018 – 2024)

El componente espacial se cierra con dos piezas complementaria: (i) la identificación de municipios “nodo” donde el rezago, espacial de formalidad (WX\_formal\_gap) alcanza los promedios más altos del período y, por lo tanto, concentra el efecto de contagio de la desigualdad; y (ii) la descomposición de la varianza del modelo con efectos aleatorios, útil para distinguir cuanto de la variación de la brecha es entre municipios, y cuanto es dentro de ellos a lo largo del tiempo. Ambas piezas robustecen la lectura sustantiva: la desigualdad salarial de género tiene anclajes territoriales reconocibles y, al mismo tiempo, muestra una dinámica idiosincrática muy intensa que exige respuestas locales con foco temporal.

En primer término, el cálculo del promedio municipal del rezago espacial de formalidad para 2018-2024 permite ordenar una lista de “hotspots”:

Tabla 4.11. Municipios con mayor brecha espacial promedio (2018–2024)

Municipio, Estado	Brecha espacial promedio
BAVIACORA, Sonora	0.75
JOSE JOAQUIN DE HERRERA, Guerrero	0.55555556
ALTO LUCERO DE GUTIERREZ BARRIOS, Veracruz	0.51428571
TELCHAC PUEBLO, Yucatán	0.45483193
SANTA MARIA IPALAPA, Oaxaca	0.35714286
MONTECRISTO DE GUERRERO, Chiapas	0.35238095
SALINA CRUZ, Oaxaca	0.30346518
MAGDALENA, Veracruz	0.29916836
SAN LUCAS, Chiapas	0.25018938
Tequila, Veracruz	0.24157211

Fuente: Elaboración propia con base en microdatos de la ENOE (2018–2024) e información geoespacial del Marco Geoestadístico INEGI 2020.

Este subconjunto, distribuido sobre todo en Sonora, Guerrero, Veracruz, Oaxaca y Chipas, revela que el efecto espacial no es aleatorio: se concentra regionalmente em zonas con estructuras laborales rígidas, densidad intencional limitada y trayectorias, históricas que dificultan la incorporación femenina a empleos formales. Dicho de otro modo, en estos municipios el entorno inmediato refuerza la brecha local porque los vecinos exhiben, de manera

persistente, ventaja masculina de formalidad que se transmiten por movilidad cotidiana, encadenamiento productivo o normas de concentración que se limitan a escala subregional.

Conviene hacer dos precisiones. Primer, el mapa de  $WX\_formal\_gap$  con estos nodos no describe “culpables”, sino puntos de palanca: lugares donde una política de formalización con enfoque de género articula con municipios colindantes, tiene retornos más altos porque rompe la transmisión territorial del sesgo. Segundo, la ausencia de datos en algunos municipios (zonas grises) se localiza en áreas rurales o de difícil acceso; esto puede reflejar tanto limitaciones de levantamiento como baja participación femenina formal. En términos operativos, el vacío informativo es en sí mismo un indicador de prioridad: si no sabemos lo suficiente, o si la formalidad femenina es marginal, la intervención requiere estadística de proximidad (trazadores locales, registros administrativos) y servicios de cuidado/movilidad que hagan viable la inserción laboral.

En segundo término, la descomposición de la varianza del modelo con efectos aleatorios confirma que al mayor parte de la dinámica es idiosincrática. El 92.8% de la variación de la brecha se explica por el componente intra-municipal (idiosyncratic), es decir, por fluctuación en el tiempo dentro de cada municipio, shocks transitorios, recomposición sectorial, cambios de reglas de concentración o medición, solo el 7.2% corresponde a diferencias estructurales entre municipios (individual).

NOTA. En la estimación se obtuvieron tablas completas de coeficientes, errores estándar y niveles de significancia para todas las especificaciones (Mincer, Heckman, descomposiciones Oaxaca–Blinder y modelos espaciales), así como gráficos adicionales de la evolución temporal de la brecha y resultados espaciales desagregados  $WX$  vs  $X$ . No obstante, su inclusión íntegra en el cuerpo de la tesis implicaría añadir un número considerable de cuadros y figuras, lo que sobrecargaría la exposición y dificultaría la lectura, sin modificar las conclusiones sustantivas del trabajo. Por esta razón, en el documento se opta por presentar únicamente los resultados sintéticos y las estimaciones más relevantes para responder a las preguntas de investigación, privilegiando la interpretación económica por encima del reporte exhaustivo de cada parámetro individual. Las tablas completas de coeficientes, los gráficos de evolución de la brecha y los resultados espaciales desagregados forman parte del archivo de salidas de Stata/R y del paquete

de replicación del estudio, y se encuentran disponibles para consulta bajo solicitud al autor o al comité de tesis.

## Conclusiones

El análisis a través de diferentes maneras de la ENOE de los años 2019-2023, estimaciones mincerinas, correcciones de sesgo y descomposición de O-B, doble y triple, además del análisis espacial mediante el modelo SLX a nivel municipal, nos dan una importante señal de que la brecha salarial por género en México es persistente, estadísticamente robusta y de base estructural. El componente no explicado domina en los tres cortes, incluso tras controlar por educación, edad, estado civil, sector, formalidad/beneficios y demás, las probabilidades de acceso a puestos con primas (industria, servicios avanzados, salarios mínimos >3, cobertura de salud), no son simétricas, por sexo. El ajuste con selección es claro, aun cuando exista salario observado, la brecha residual permanece. Mediante la descomposición triple los retornos, explican la mayor fracción, a su vez la interacción (I) prevalece con importancia al dominar mediante dotaciones favorables a varones, donde se encuentran con precios que los benefician, finalmente, en SLX a identificado a la formalidad como bisagra local, y vehículo de derrame, vía  $WX\_formal\_gap$ : es decir, la formalidad como un factor elemental local, dentro del mismo municipio, al mismo tiempo como un canal de transmisión global.

En los municipios donde los hombres concentran el empleo formal, la brecha salarial tiende a ser mayor, ese patrón se replica hacia los municipios vecinos a través de las interconexiones laborales y geográficas generando efectos de derrame espacial de la desigualdad.

Incluso con dotaciones similares, los retornos divergen, la idea de organizaciones generalizadas (Acker, 1990) sugiere que reglas aparentemente neutras (disponibilidad horaria, movilidad, entre otras), codifican expectativas de género. En esa clave, (Ridgway, 2011) conceptualiza el género como sistema de estatus: es decir, estereotipos, en su mayoría sutiles, y por naturaleza, generan premios o credenciales específicas por géneros, en su caso descuentos a las feminizadas, otros autores como (Kanter, 1977) describe dos mecanismos que vemos reproducidos en los datos; “techo de cristal” una subrepresentación en la cúspide de los puestos laborales con primas, y los *efectos de token* (evaluación hipervisibilizada de minorías en roles dominado por el otro sexo. Complementariamente (Williams, 1992) postula que el “escalador de cristal”: en relación con ocupaciones feminizadas, es donde los hombres ascienden más rápido, esto está presente en nuestras estimaciones, a través de la

subrepresentación en puesto con primas, >3 SM, industria formal, servicios con carrera, ya que cuando se llega a acceder aún con las bajas probabilidades.

Más evidencia como la penalización por maternidad (Correl, Bernard, & Paik, 2007; Budig & Hodges, 2010) nos aportan sustento en coeficientes asociados a estado civil o hijos, ya que la “madre” tiende a ser percibida como menos comprometidas, en tanto la “paternidad” puede operar, paradójicamente, como señal positiva, en nuestros datos, el patrón de penalizaciones femeninas asociadas a la carga de cuidados resulta consistente, incluso controlado por sector y educación: como mecanismos de piso pegajoso.

La antropología económica ha insistido en que los arreglos productivos no son meramente decisiones individuales, están sujetos a instituciones y normas sociales y locales. Autores como (Folbre, 2001; Elson, 2017; Fraser, 2016), en materia de economía del cuidado, y de reproducción social, hacen énfasis en como la organización fuera del mercado condiciona trayectorias dentro del mercado, en los resultados espaciales, se observa como los municipios con baja formalidad femenina encadenan inserciones precarias, alta rotación, y ausencia de trayectorias de ascenso, cuando, además la vecindad exhibe brechas de formalidad masculinas altas, el SLX detecta derrame: pues no solo las instituciones locales, sino las prácticas de contratación y redes de los municipios cercanos difunden desventajas de acceso. Estos aportes son importantes para el análisis de periferias rurales, pueblos indígenas y corredores migratorios, donde las restricciones locales forman parte del acceso a empleos formales.

La economía feminista ha mostrado que ocupaciones feminizadas sufren devaluación retributiva: mismo nivel de habilidad, pero menor paga, caracterizado por sesgos históricos, y menos poder de negociación, está presente en la descomposición triple, la C de coeficientes captura precios distintos para idénticas credenciales. Los resultados recientes sobre habilidades y tareas en América Latina (Bonavida & Gasparini, 2024) plantean que la matriz de tareas se remunera de manera heterogénea, y que las mujeres se concentran en tareas subpagadas, pese a niveles educativos similares.

En nuestros resultados, se indica que las primas por servicios médicos, es alta para ambos y a veces mayor para las mujeres, esto sugiere que, a precio igual, ellas ganan, sin embargo, su probabilidad de acceso a puestos es menos, con lo que la contribución al promedio femenino

sigue siendo pequeña, se puede ver otra perspectiva del techo de cristal donde que necesariamente es la ausencia de talento, sino un filtro al acceso y valoración de tareas.

Perspectivas neoclásicas tienden a limitarse a los datos y a la ocupación neta, una perspectiva no feminista como (Becker G. , 1993; Polachek, 1981) enfatiza que hay una especialización del hogar, capital humano y elección ocupacional, es decir la diferencia reflejan una elección por preferencias y costos, como un horario mas flexible. (Hakim, 2000) postula que la brecha responde a elecciones heterogéneas entre “orientación a la carrera y orientación al hogar. Sin embargo, en nuestros resultados por Heckman se muestra que quien participa si importa, pero no a la brecha condicionada, persisten los retornos asimétricos aun cuando se esté dentro de la ecuación.

Más en SLX, y los efectos de vecindad en la formalidad: existe un rezago espacial robusto, (estructura conformada por instituciones, redes y reglas) que multiplican elecciones individuales, es decir que existe una preferencia individual restrictiva, otro punto es en los umbrales altos del salario, ( $>3SM$ ) existen para mujeres que llegan a la cúpula; sin embargo, no es la falta de ambición, si no la tasa de acceso, incluso la capacidad de negociación. Los aportes neoclásicos son pertinentes al evaluar las trayectorias individuales, pero no alcanzan a explicar la magnitud de componentes como C y de fenómenos de tipo derrame territorial.

El hallazgo espacial es clave:  $WX\_formal\_gap$  positivo indica que la desigualdad se propaga por redes de vecindad. La geografía económica (Glaeser, 2011; Moretti, 2012; Storper, 2013), ha mostrado que las aglomeraciones elevan productividad y salarios, pero también segmentan: cuando la puerta formal se abre más para varones, las cadenas de valor y rutinas intermunicipales replican esa preferencia, en sociología organizacional, (DiMaggio & Powell, 1983) hablaron de isomorfismo: las organizaciones imitan prácticas de sus pares; nuestro SLX sugiere una versión territorial de ese proceso. Si varios municipios vecinos premian perfiles masculinos y exigen disponibilidad incompatible con cuidados, la vecindad se convierte en multiplicador de la brecha. Por eso, la política no puede ser solo “macro” ni solo “empresa por empresa”: debe tener una escala regional (corredores, clústeres, áreas metropolitanas).

Es importante no limitarnos únicamente en los datos, bien lo explica la descomposición O-B donde los valores no explicados son un fuerte peso de la brecha, la interseccionalidad

(Crenshaw, 1989; Collins, 2022) nos recuerda que el género no opera solo, en México, factores como territorio, clase, condición étnico-lingüística y estatus migratorio, se cruzan con el sexo, a través de los coeficientes de ruralidad, algunos de baja escolaridad y subocupación, muestran un efecto de piso pegajoso, mayormente denso para mujeres indígenas rurales o en periferias urbanas. La suma de estas características adhiere a penalizaciones ya que están altamente vinculadas a más características de baja retribución como empleos sin prestaciones, rotaciones y tiempos muertos en traslados.

Este trabajo de tesis ofrece una amplia evidencia de la estructura social, puede ser importante en el análisis de igualdad precios, con las implicaciones como una transparencia retributiva, auditorías salariales periódicas, por sexo y ocupación, tabuladores con bandas y criterios explícitos y sostenidos de promoción, como mejorar las cláusulas de corrección salarial cuando la parte no explicada supere umbrales predefinidos. De igual manera crear incentivos a la afiliación, así como conciliación y cuidados, (licencias parentales simétricas, horarios flexibles y bien pagados transporte seguro, y estancias extendidas), que reduzcan la penalización por tiempo, también coordinar política intermunicipal, donde SLX detecta derrames de formalidad: convenios de corredor con empresas ancla para homogeneizar estándares de contratación, compras públicas con criterios de igualdad salarial e intervención de “clusters” en zonas críticas, (guarderías, movilidad laboral, bolsa de trabajo compartida).

Además, este trabajo aporta un abanico de opciones de investigación, que va desde replicar metodológicamente para diferentes cortes temporales, así como una amplia gama de información reproducible con datos de la ENOE con estudios comparables, períodos, ramas e incluso otras encuestas que compartan relación metodológica y temática, como evaluar la transparencia salarial, guarderías, licencias y formalización femenina como mecanismo de acceso; para investigar el incremento de la probabilidad del empleo formal femenino, determinantes para traspasar el techo de cristal, evaluación de las tareas y habilidades como matriz de tareas, aglomeraciones y empleo ancla, para identificar el efecto de la llegada de empresas, modelos de duración y carreras para identificar las transiciones tanto formales como informales, dependencia espacial dinámica explorar los mecanismos de derrames espaciales, modelos machine learning para identificar perfiles femeninos con mayor probabilidad a ganar más, entre otros. Una de las características importantes de nuestra base

de datos nace de la diversa manera de armonizar con más fuentes, y otras encuestas como con ENUT, ENSANUD, ENADID, ENIGH, DENUE, IMSS, censos, entre otros.

## **Bibliografía**

### **Trabajos citados**

- Acker, J. (1990). Jerarquías, puestos de trabajo, cuerpos: Una teoría de las organizaciones con perspectiva de género. *Género y sociedad*, 4(2), 139-158.
- Albuja, W., y Enríquez, M. J. (2018). Análisis de la discriminación laboral hacia las mujeres en Ecuador 2017-2016. *Convergencia Revista de ciencias sociales*. (78), 13-41. doi:<https://doi.org/10.29101/crcs.v0i78.9692>
- Andres-Rosales, R., Czarnecki, L., y Mendoza, G. (2019). Un análisis espacial de la precariedad y la brecha salarial de género en México, 2005-2018. *The Journal of Chinese Sociology*, 6(1), 13.
- Anselin, L. (2022). Econometría espacial. Manual de análisis espacial en las ciencias sociales.
- Arceo, E., y Raymundo, C. (2014). Evolución de la brecha salarial de género. *Trimestre Económico*, 81(323), 619-653. Obtenido de [https://www.scielo.org.mx/scielo.php?pid=S2448-718X2014000300619yscript=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S2448-718X2014000300619yscript=sci_arttext)
- Arceo-Gómez, E., y Torres, P. (2021). Brechas salariales por autoidentificación indígena y rasgos lingüísticos en México. *Sobre México Temas de Economía*, 1(3), 129-161. doi:<https://doi.org/10.48102/rsm.vi3.94>
- Babcock, L., Laschever, S., Gelfand, M., y Small, D. (2003). Nice girls don't ask. Harvard business review. 81(10), 14-14.
- Becker, G. (1993). *Un tratado sobre la familia: Edición ampliada*. Harvard University Press.
- Becker, H. (1971). *Trabajo sociológico*. Editorial Transaction.
- Blau, F., y Khan, L. (2017). La brecha salarial de género: alcance, tendencias y explicaciones. *Revista de literatura económica*, 55(3), 789-865.
- Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 436-455. doi:<https://doi.org/10.2307/144855>.
- Bonavida, C., y Gasparini, L. (2024). El rol de las habilidades y las tareas de las brechas de género en América Latina. *Documentos de Trabajo del CEDLAS*, 1-57. doi:<http://sedici.unlp.edu.ar/handle/10915/166454>.

- Budig, M., y Hodges, M. (2010). Diferencias en desventaja: Variación en la penalización por maternidad según la distribución de ingresos de las mujeres blancas. *American Sociological Review*, 75(5), 705-728.
- Bursal, M., y Sentürk, I. (2023). Determinants of gender discrimination in working hours: the Blinder Oaxaca decomposition method. *MKU Journal of Social and Economics*, 1014-1034. Obtenido de <https://dergipark.org.tr/en/pub/kmusekad/issue/81697/1341968>
- Buser, T., Geijtenbeek, L., y Plug, E. (2018). Sexual orientation, competitiveness and income. *Elsevier*. doi:<https://doi.org/10.1016/j.jebo.2018.03.017>
- Castro Lugo, D., Rodríguez, P., y Ramos Lobo, R. (2021). ¿ Existe convergencia regional en la brecha salarial por género en México? *Economía, sociedad y territorio*, 21(67), 685-717.
- Collins, P. (2022). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge.
- Correl, S., Bernard, S., y Paik, I. (2007). Conseguir un empleo: ¿Existe una penalización por ser madre? *Revista Americana de Sociología*, 112(5), 1297-1338.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1, 139-167.
- de Castro Romero, L., Barroso, V., y S., S. (2025). Brecha salarial de género en la gestión hotelera: La evidencia en Europa. *Annals of Tourism Research Empirical Insights*, 6(1), 100-169.
- DiMaggio, P., y Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147-160.
- Elhorst, J. (2003). Especificación y estimación de modelos de datos de panel espaciales. *Revista científica regional internacional*, 26(3), 244-268.
- Elson, D. (2017). Reconocer, reducir y restringir el trabajo de cuidados no remunerado: cómo cerrar la brecha de género. *New Labor Forum*, 26(2), 53-61.
- Folbre, N. (2001). *El corazón invisible: Economía y valores familiares*. New Press, Nueva York.
- Fortin, N. L. (2011). Métodos de descomposición en economía. *Handbook of labor economics*, 4, 1-102. doi:[https://doi.org/10.1016/S0169-7218\(11\)00407-2](https://doi.org/10.1016/S0169-7218(11)00407-2)
- Fraser, N. (2016). Contradicciones del capital y el cuidado.
- Glaeser, E. (2011). *El triunfo de la ciudad: cómo los espacios urbanos nos humanizan*. Pan Macmillan.

- Hakim, C. (2000). *Opciones de estilo de vida y trabajo en el siglo XXI: teoría de las preferencias*. . Oxford University Press.
- Heckman, J. (1994). Sesgo de selección de muestra como error de especificación. *Econometría: Revista de la Sociedad Econométrica*, 47(1), 153-161. doi:<https://doi.org/10.2307/1912352>
- Hedija, V. (2023). Brecha salarial de género en la República Checa: enfoque en la gestión. *Investigación económica-Ekonomska istraživanja*, 36(3). doi:<https://doi.org/10.1080/1331677X.2023.2263510>
- Hernández Ballesteros, V., y Almonte, L. (2024). Pobreza y especialización productiva. Un estudio de econometría espacial para los municipios de Oaxaca, México, 2020. *Revista de economía*, 41(103), 99-127.
- Instituto Nacional de Estadística y Geografía. (s.f.). *Encuesta Nacional de Ocupación y Empleo (ENOE)*. Obtenido de INEGI: <https://www.inegi.org.mx/programas/enoe/15ymas/>
- Jann, B. (2008). Descomposición de Blinder-Oaxaca para modelos de regresión lineal . *The Stata Journal* , 8(4), 453-479. doi:<https://doi.org/10.1177/1536867X0800800401>
- Kanter, R. (1977). Hombres y mujeres de la corporación. *Nueva York*, 209.
- Killingsworth, M. (1983). Brechas salariales entre sindicatos y no sindicatos y aumentos salariales: Nuevas estimaciones de un estudio transversal de la industria. *The Review of Economics and Statistics*, 332-336.
- Lombardia, M., Lopez-Vizcaino, E., y Rueda, C. (2022). Un nuevo enfoque para la descomposición de la brecha salarial de género por actividad económica. *Journal of the Royal Statistical Society, Serie A: Statistics in Society*, 185(1), 219-245.
- Luz Tovar, C., y Pineda Salazar, S. I. (2023). Análisis regional de la discriminación salarial de los trabajadores indígenas en México. *Nósis. Revista de ciencias sociales*, 32(64), 47-72. doi:<https://doi.org/10.20983/noesis.2023.2.3>
- Matuszewka-Janica, A. (2025). Autocorrelación espacial del indicador de brecha salarial de género en las macrorregiones de la Unión Europea. *Folia Oeconomica Stetinensia*, 25(1), 180-200.
- Mincer, J. (1974). La función de ingresos del capital humano. *Escolaridad, experiencia e ingresos*, 83-96.
- Moretti, E. (2012). *La nueva geografía del empleo*. Houghton Mifflin Harcourt.
- Oaxaca, R. (1973). Diferenciales salariales entre hombres y mujeres en los mercados laborales urbanos. *Revista económica internacional*, 694-709.

- Oaxaca, R., y Ransom, M. (1994). Sobre la discriminación y la descomposición de los diferenciales salariales. *Journal of Econometrics*, 61(1), 5-21.  
doi:[https://doi.org/10.1016/0304-4076\(94\)90074-4](https://doi.org/10.1016/0304-4076(94)90074-4)
- Otsu, T., y Tanaka, S. (2022). Empirical likelihood inference for Oaxaca-Blinder decomposition. *Economics letters*(219).  
doi:<https://doi.org/10.1016/j.econlet.2022.110812>
- Polachek, S. (1981). Autoselección ocupacional: Un enfoque de capital humano para las diferencias de género en la estructura ocupacional. *The Review of Economics and Statistics*, 60-69.
- Ridgway, C. (2011). *Enmarcado por el género: Cómo persiste la desigualdad de género en el mundo moderno*. Oxford University Press.
- Rodríguez Pérez, R. E. (2014). Discriminación salarial de la mujer en el mercado laboral de México y sus regiones. *Economía, sociedad y territorio*, 14(46), 687-714. Obtenido de [https://www.scielo.org.mx/scielo.php?pid=S1405-84212014000300004&script=sci\\_abstract&lng=pt](https://www.scielo.org.mx/scielo.php?pid=S1405-84212014000300004&script=sci_abstract&lng=pt)
- Rodríguez Pérez, R. E. (2019). Diferencia salarial por género entre el sector público y privado formal-informal en México. *Revista de Economía*, 36(93), 62-89.  
doi:<https://doi.org/10.33937/reveco.2019.108>
- Rodríguez Pérez, R. E., y Limas Hernández, M. (2017). El análisis de las diferencias salariales por género por áreas profesionales en México, abordado desde un enfoque regional, 2015. *Estudios Sociales*, 27(49), 121-150. Obtenido de [https://www.scielo.org.mx/scielo.php?pid=S0188-45572017000100121&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0188-45572017000100121&script=sci_arttext)
- Romero, M., y Prudencio, V. (2022). Brecha salarial de género: análisis exploratorio con la ENOE, 2013 y 2018.
- Rosales, R., Mendoza, G., y R., Q. (2023). Análisis espacial de la desigualdad salarial y la precariedad laboral en el sector turístico mexicano. *Contaduría y administración*, 68(2).
- Salce Díaz, F. (2021). Evolución y análisis de la discriminación salarial por género en Chile. *Trimestre Económico*, 88(349), 39-75.  
doi:<https://doi.org/10.20430/ete.v88i349.984>
- Scotti, C. M., y Manzano, J. R. (2015). Análisis de la brecha de ingresos entre las personas con y sin discapacidad en ocupaciones no calificadas. *Revista Latinoamericana de Población*, 9(16), 35-58. doi:<https://doi.org/10.31406/relap2015.v9.i1.n16.2>
- Smith, K., y R., G. (2013). Explorando la penalización salarial espacial para las mujeres: ¿Importa dónde se vive? *Investigación en Ciencias Sociales*, 42(5), 1390-1401.

- Storper, M. (2013). *Claves para la ciudad: Cómo la economía, las instituciones, la interacción social y la política influyen en el desarrollo*. Princeton University Press.
- Valdivia, L. P. (2011). Segmentación laboral, educación y desigualdad salarial en México. *Revista mexicana de sociología*, 73(1), 139-175. Obtenido de [https://www.scielo.org.mx/scielo.php?pid=S0188-25032011000100005&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S0188-25032011000100005&script=sci_arttext)
- Williams, C. (1992). La escalera mecánica de cristal: Ventajas ocultas para los hombres en las profesiones <<femeninas>>. *Problemas sociales*.
- Yusuf, F., San Sebastián, M., y Vaezghasemi, M. (2023). Explaining gender inequalities in overweight people: a Blinder-Oaxaca decomposition. *International Journal for Equity in Health*, 22(1), 1-7. doi:10.1186/s12939-023-01973-9

### **Bibliografía complementaria**

- Aguilar-Rodríguez, A., Miranda, A., y Zhu, Y. (2018). Descomponiendo la brecha salarial lingüística entre las minorías étnicas indígenas de México: ¿se debe todo a lo observable? *Boletín de Economía*, 38, 689–695.
- Arceo, E., y Raymundo, C. (2014). Evolución de la brecha salarial de género. *Trimestre Económico*, 81(323), 619–653.
- Arpi Mayta, R., y Arpi Quilca, L. (2018). Desigualdad del ingreso laboral y nivel educativo entre grupos étnicos en Perú. *Comunicación*, 9(1), 56–67.
- Benitez, D., y Espinoza, B. (2018). Discriminación salarial por género en el sector formal en Ecuador usando registros administrativos. Quito: INEC.
- Bermúdez, K. J., y Cota, J. E. (2009). Discriminación salarial por género en México. *Revista latinoamericana de Economía*, 40(156), 77–100.
- Canavire-Bacarreza, G. U. (2016). Informality and Mobility in the Labor Market: A pseudopanel's approach. *LAJED*, 14(27), 57–76.
- Canedo, A. (2019). Labor market discrimination against indigenous peoples in Mexico: A decomposition analysis of wage differentials. *Iberoamericana–Nordic Journal of Latin American and Caribbean Studies*, 48(1).

- Cepeda, L., Cardona, L., y Barón, J. D. (2014). La brecha salarial de género entre los universitarios recién titulados: ¿Qué tanto influye el tipo de carrera universitaria? *El Trimestre Económico*, 441–477.
- Choe, C., Jung, S., y Oaxaca, R. (2022). What's the Risk from Competing? Competition Aversion and the Gender Wage Gap. IZA DP 15048.
- Christian, D. I., y Sibyl, P. (2023). Análisis regional de la discriminación salarial de los trabajadores indígenas en México. *Noésis. Revista de Ciencias Sociales y Humanidades*, 47–72.
- Fairlie, R. (1999). La ausencia de empresas propiedad de afroamericanos: un análisis de la dinámica del trabajo por cuenta propia. *Journal of Labor Economics*, 17(1), 80–108.
- Fuentes, J., Palma, A., y Montero, R. (2005). Discriminación por género en Chile, una mirada. *Estudios de Economía*, 32(2), 133.
- Groisman, F. (2014). Empleo, salarios y desigualdad en Argentina: Análisis de los determinantes distributivos. *Problemas del Desarrollo*, 45(177), 59–86.
- Hernández, C., Vásquez, P., y Narváez, N. (2010). La reducción salarial de Oaxaca-Blinder: Métodos, críticas y aplicaciones. *Revista de Economía del Caribe*, 5.
- Hlavac, M. (2014). Oaxaca: Descomposición de Blinder-Oaxaca en R. SSRN.
- Jung, S. (2017). The gender wage gap and sample selection via risk attitudes. *International Journal of Manpower*, 38(2), 318–335.
- Kolpashnikova, K., y Kan, M. Y. (2020). The gender gap in the United States. *Demographic Research*, 43, 1067–1080.
- Kröger, H., y Hartmann, J. (2021). Extensión del método de descomposición Kitagawa-Oaxaca-Blinder a datos de panel. *The Stata Journal*, 21(2), 360–410.
- Lamprea Barragán, T. C. (2019). Brechas salariales de género en el mercado laboral urbano colombiano: efectos del sector económico.
- Linthon-Delgado, D. E.-H. (2022). Descomposición de la brecha salarial de género en el Ecuador. *Revista Mexicana de Economía y Finanzas*, 17(1).
- Moreno Soler, M. P. (2021). La brecha rosa: aplicación del modelo de Oaxaca-Blinder para comprender la diferencia en el ingreso entre la mujer urbana y rural en Colombia.

- Mussa, R. (2014). Extending the Oaxaca–Blinder Decomposition to the Independent Double Hurdle Model: With Application to Parental Spending on Education in Malawi. *Studies in Economics and Econometrics*, 38(3), 39–54.
- Ñopo, H. (2008). Una extensión de la descomposición de Blinder–Oaxaca a un continuo de grupos de comparación. *Economics Letters*, 100(2), 292–296.
- Ospino, C., Roldan, V., y Barranza, N. (2010). Oaxaca–Blinder wage decomposition: Methods, critiques and applications. A literature review. *Revista de Economía del Caribe*, 5, 237–274.
- Otero-Cortés, A., y Acosta-Ariza, E. (2022). Desigualdades en el mercado laboral urbano–rural en Colombia, 2010–2019. *CS SPE1*, 173–219.
- Payán, L. D. (2015). Las brechas de género en la calidad del empleo en México. *Análisis Económico*, 30(75), 89–112.
- Quiñones, D. M., y Rodríguez, J. A. (2011). Rendimiento de la educación en las regiones colombianas: un análisis usando la descomposición Oaxaca–Blinder. *Sociedad y Economía*, 1(20), 38–68.
- R., O. (1973). Male–Female Wage Differentials in Urban Labor Markets. *Revista Economica Internacional*, 14(3), 693–709.
- Simon, H. (2006). Diferencias salariales entre hombres y mujeres en España: una comparación internacional con datos. *Investigaciones Económicas*, 30(1), 55–87.
- Teixeira de Matos Vázquez, C. (2021). Género y mercado laboral en España. TFG, Universidad de Valladolid.

## Anexo A. Descriptor de las variables

- Variable dependiente
  - *lsalario*: Logaritmo natural del ingreso mensual reportado
- **Variables explicativas**
  - *ur*: Dicotómica para zonas urbanas
  - *edad*: Edad
  - *grupo\_edad*: Categoría de edad (menor de 14 años, menor a 24, menor a 34, menor a 44, menor a 54, menor a 64, mayor a 64 años)
  - *e\_civil*: Estado civil: considera “soltero = 0” personas, solteros divorciados, y viudos), y “casados = 1” personas en unión libre y casados.
  - *grado\_escolar*: Último grado escolar finalizado (Menor a primaria; preescolar o ninguno, Primaria, Secundaria, Preparatoria; bachillerato técnico y general, Profesionista; Normal, Licenciatura, Maestría y Doctorado.
  - *años\_esc*: Hasta qué año aprobó en la escuela (en cantidad de años, 1, 2, 3, 4, ... etc)
  - *seg\_soc*: población que cuenta con una condición de acceso a instituciones de seguridad social
  - *rama*: Clasificación de la población ocupada por sectores de actividad económica (Construcción, Industria manufacturera, Comercio, Servicios, Otros y Agropecuario).
  - *rama\_est1*: clasificación de la población ocupada por sector de actividad económica (Primario, Secundario, Terciario)
  - *Sabe\_leer\_escribir*: Sabe leer o escribir un recado
- Variables de control adicionales.
  - *n\_hijos3*: Tiene de 1 a 3 hijos
  - *n\_hijos6*: Tiene, más de 4 o más hijos
  - *ing7c*: Clasificación de la población ocupada por nivel de ingreso.

## Anexo B. Manual de construcción de la base de datos ENOE

Consideraciones técnicas; para compilar una base de datos anual, ENOE contempla 5 cuestionarios, (Hogar, Vivienda, Socioeconómico, Cuestionario 1 y Cuestionario 2), solo para un trimestre, por lo tanto, si usted desea hacer el compilado de ENOE, para el primer trimestre “ENOE\_t1” deberá unir mediante la metodología de “merge” estos cuestionarios, mediante la sugerencia de la ficha técnica de INEGI

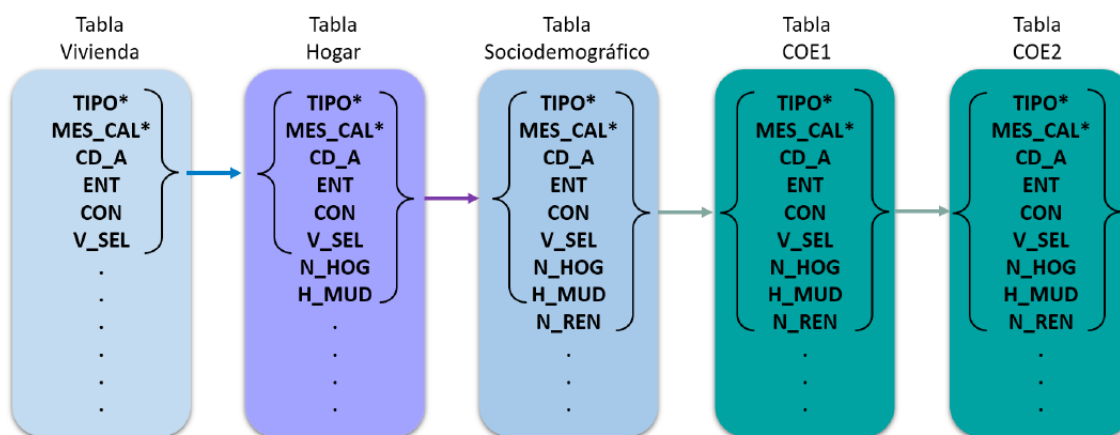


Ilustración 1. Nota: La figura muestra la relación entre los cinco cuestionarios que integran la Encuesta Nacional de Ocupación y Empleo (ENOE). Cada archivo (Vivienda, Hogar, Sociodemográfico, COE1 y COE2) debe vincularse mediante las variables comunes TIPO, MES\_CAL, CD\_A, ENT, CON, V\_SEL, N\_HOG, H\_MUD y N\_REN, siguiendo la metodología de merge recomendada por el INEGI para la construcción de bases trimestrales. Fuente: Instituto Nacional de Estadística y Geografía (INEGI) (2024). Conociendo la base de datos de la ENOE. Dirección General de Estadísticas Sociodemográficas. Disponible en <https://www.inegi.org.mx/programas/enoe/>

Sin embargo, si se requiere generar la base de datos anual deberás unir los 4 trimestres correspondientes, mediante la metodología “append”, a continuación, se sugiere una metodología para conformar la base de datos 2024, esta forma se puede aplicar en los demás años, a consideración cambios técnicos que se susciten, como la conformación de nuevas variables (preguntas) así como la eliminación, en la actualización de cada cuestionario. Esta línea de código aplica para el Software estadístico Stata 17. Se omiten referencias técnicas del lenguaje, sin embargo, se aportan comentarios con referencia a cuestiones técnicas de dudosa resolución, tome en cuenta que la ubicación que se expresa para guardar, exportar e importar las bases de datos, edítelas a conveniencia.

Consideraciones previas: en la página de INEGI, rediríjase a la parte de microdatos ENOE, la cual para el software de Stata requiere los archivos “.dta” una vez descargados podrás observar que tienes 5 bases de datos formato “.dta” como estos: por ejemplo, estos son del trimestre 1 del año 2024.

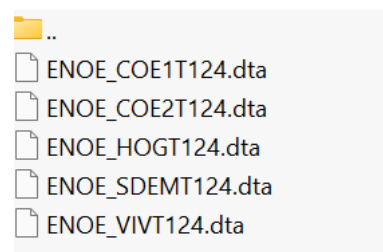


Ilustración 2. Referencia de descargas, Archivos base de la ENOE correspondientes al primer trimestre de 2024

```
*****AÑO 2024*****
```

```
***** para el Trimestre 1 del 2024 *****
```

```
cd "E:\Usuarios\783251\Downloads\enoe_2024_trim1_dta"
us ENOE_VIVT124.dta
us ENOE_HOGT124.dta, clear
merge m:m mes_cal cd_a est using ENOE_VIVT124.dta
drop _merge
```

\*Guarde actualización de base se sugiere crear una carpeta específica para estas bases que servirán como soporte por ejemplo cree una carpeta llamada "Merges" \*

```
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t1_VH.dta"
***
```

```
merge m:m mes_cal cd_a est using ENOE_SDEMT124.dta
drop _merge
```

\*Para guardarlos en la carpeta Merges bases

```
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t1_VHS.dta"
***
```

```
merge m:m mes_cal cd_a ent using ENOE_COE1T124.dta
drop _merge
```

\*Para guardarlos en la carpeta Merges bases

```
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t1_VHSC1.dta"
***
```

```
merge m:m mes_cal cd_a ent using ENOE_COE2T124.dta
drop _merge
```

```
save "E:\Usuarios\783251\Desktop\Steve Economia\Tesis\Datos\bases
trimestrales\2024_T1.dta"
***
```

```
clear
```

\*\*\* En este punto usted ha podido conformar la base de datos 2024 correspondiente al primer trimestre "2024\_T1.dta" esta última

conviene guardarla en otra carpeta específica para guardar las bases creadas trimestralmente\*\*\*

\*\*\*\*\* para el Trimestre 2 del 2024 \*\*\*\*\*

```
cd "E:\Usuarios\783251\Downloads\enoe_2024_trim2_dta"
us ENOE_VIVT224.dta
us ENOE_HOGT224.dta, clear
merge m:m mes_cal cd_a est using ENOE_VIVT224.dta
drop _merge
*Para guardarlo en la carpeta de Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t2_VH.dta"
***
merge m:m mes_cal cd_a est using ENOE_SDEMT224.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t2_VHS.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE1T224.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t2_VHSC1.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE2T224.dta
drop _merge
save "E:\Usuarios\783251\Desktop\Steve Economia\Tesis\Datos\bases
trimestrales\2024_T2.dta"
***
clear
```

\*\*\*\*\* para el Trimestre 3 del 2024 \*\*\*\*\*

```
cd "E:\Usuarios\783251\Downloads\enoe_2024_trim3_dta"
us ENOE_VIVT324.dta
us ENOE_HOGT324.dta, clear
merge m:m mes_cal cd_a est using ENOE_VIVT324.dta
drop _merge
*Para guardarlo en la carpeta de Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t3_VH.dta"
***
merge m:m mes_cal cd_a est using ENOE_SDEMT324.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
```

```

save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t3_VHS.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE1T324.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t3_VHSC1.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE2T324.dta
drop _merge
save "E:\Usuarios\783251\Desktop\Steve Economia\Tesis\Datos\bases
trimestrales\2024_T3.dta"
***
clear

```

\*\*\*\*\*Para el trimestre 4 del 2024\*\*\*\*\*

```

cd "E:\Usuarios\783251\Downloads\enoe_2024_trim4_dta"
us ENOE_VIVT424.dta
us ENOE_HOGT424.dta, clear
merge m:m mes_cal cd_a est using ENOE_VIVT424.dta
drop _merge
*Para guardarlo en la carpeta de Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t4_VH.dta"
***
merge m:m mes_cal cd_a est using ENOE_SDEMT424.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t4_VHS.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE1T424.dta
drop _merge
*Para guardarlos en la carpeta Merges bases
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\t4_VHSC1.dta"
***
merge m:m mes_cal cd_a ent using ENOE_COE2T424.dta
drop _merge
save "E:\Usuarios\783251\Desktop\Steve Economia\Tesis\Datos\bases
trimestrales\2024_T4.dta"
***
clear

```

\*\*\* Hasta aquí usted ha conformado las bases de datos del 2024, es decir tiene, cuatro bases de datos correspondientes al trimestre; I,

II, III y IV \*\*\* ahora requiere unir las 4 mediante la metodología append \*\*\*

\*;Consideraciones técnicas!

Algunas variables están guardadas bajo forma string que impide la unión correcta de las bases, estas las sugiera eliminar mediante el comando: esto está presente en diferentes trimestres de diferente año, el software las señala y pueden cambiar, se sugiere eliminarlas porque su aporte no resulta útil. Su importancia será a a su consideración o puede modificarlas mediante el comando "destring" otro momento donde puede las variables tipo string puede ocasionar problemas es en la parte de unión por trimestres bajo la metodología append, puede usar las mismas sugerencias \* drop t\_loc cs\_p14\_c cs\_p20\_des cs\_p22\_des

\*Las bases de datos trimestre 3 y 4 del 2019 tienen las variables en mayúsculas lo cual impide hacer un adecuado append en la conformación de una base\*

\* la solución a este problema es manual, es decir tienes que renombrar las variables, a continuación, dejó el listado\*

```

rename LOC loc
rename MUN mun
rename EST est
rename EST_D est_d
rename AGEB ageb
rename T_LOC t_loc
rename CD_A cd_a
rename ENT ent
rename CON con
rename UPM upm
rename D_SEM d_sem
rename N_PRO_VIV n_pro_viv
rename V_SEL v_sel
rename N_ENT n_ent
rename PER per
rename P1 p1
rename P2 p2
rename P3 p3
rename UR ur
rename FAC fac
rename R_DEF r_def
rename N_HOG n_hog
rename H_MUD h_mud
rename N_REN n_ren
rename C_RES c_res
rename PAR_C par_c
rename SEX sex
rename EDA eda
rename NAC_DIA nac_dia
rename NAC_MES nac_mes
rename NAC_ANIO nac_anio
rename L_NAC_C l_nac_c
rename CS_P12 cs_p12
rename CS_P13_1 cs_p13_1
rename CS_P13_2 cs_p13_2
rename CS_P15 cs_p15
rename CS_P16 cs_p16
rename CS_P17 cs_p17
rename N_HIJ n_hij
rename E_CON e_con
rename CS_AD_MOT cs_ad_mot
rename CS_AD_DES cs_ad_des
rename CS_NR_MOT cs_nr_mot
rename CS_NR_ORI cs_nr_ori
rename ZONA zona
rename SALARIO salario
rename CLASE1 clase1
rename CLASE2 clase2
rename CLASE3 clase3
rename POS_OCU pos_ocu
rename SEG_SOC seg_soc
rename RAMA rama
rename C_OCU11C c_ocu11c
rename ING7C ing7c
rename DUR9C dur9c
rename EMPLE7C emple7c
rename MEDICA5C medica5c
rename BUSCAR5C buscar5c
rename RAMA_EST1 rama_est1
rename RAMA_EST2 rama_est2
rename DUR_EST dur_est
rename AMBITO1 ambito1
rename AMBITO2 ambito2
rename TUE1 tue1
rename TUE2 tue2
rename TUE3 tue3
rename BUSQUEDA busqueda
rename D_ANT_LAB d_ant_lab
rename D_CEXP_EST d_cexp_est
rename DUR_DES dur_des
rename SUB_O sub_o
rename S_CLASIFI s_clasifi
rename REMUNE2C remune2c
rename PRE_ASA pre_asa
rename TIP_CON tip_con
rename DISPO dispo
rename NODISPO nodispo
rename C_INAC5C c_inac5c
rename PNEA_EST pnea_est
rename NIV_INS niv_ins
rename EDA5C eda5c
rename EDA7C eda7c
rename EDA12C eda12c
rename EDA19C eda19c
rename HIJ5C hij5c
rename DOMESTICO domestico
rename ANIOS_ESC anios_esc
rename HRSOCUP hrsocup
rename INGOCUP ingocup
rename ING_X_HRS ing_x_hrs
rename TPG_P8A tpg_p8a
rename TCCO tcco
rename CP_ANOC cp_anoc
rename IMSSISSSTE imssissste
rename MA48ME1SM ma48me1sm
rename P14APOYOS p14apoyos

```

```

rename SCIAN scian
rename T_TRA t_tra
rename EMP_PPAL emp_ppal
rename TUE_PPAL tue_ppal
rename TRANS_PPAL trans_ppal
rename MH_FIL2 mh_fil2
rename MH_COL mh_col
rename SEC_INS sec_ins
rename N_INF n_inf
rename P1A1 p1a1
rename P1A2 p1a2
rename P1A3 p1a3
rename P1B p1b
rename P1C p1c
rename P1D p1d
rename P1E p1e
rename P2_1 p2_1
rename P2_2 p2_2
rename P2_3 p2_3
rename P2_4 p2_4
rename P2_9 p2_9
rename P2A_DIA p2a_dia
rename P2A_SEM p2a_sem
rename P2A_MES p2a_mes
rename P2A_ANIO p2a_anio
rename P2B_DIA p2b_dia
rename P2B_SEM p2b_sem
rename P2B_MES p2b_mes
rename P2B_ANIO p2b_anio
rename P2B p2b
rename P2C p2c
rename P2D1 p2d1
rename P2D2 p2d2
rename P2D3 p2d3
rename P2D4 p2d4
rename P2D5 p2d5
rename P2D6 p2d6
rename P2D7 p2d7
rename P2D8 p2d8
rename P2D9 p2d9
rename P2D10 p2d10
rename P2D11 p2d11
rename P2D99 p2d99
rename P2E p2e
rename P2F p2f
rename P2G1 p2g1
rename P2G2 p2g2
rename P2H1 p2h1
rename P2H2 p2h2
rename P2H3 p2h3
rename P2H4 p2h4
rename P2H9 p2h9
rename P2I p2i
rename P2J p2j
rename P2K_ANIO p2k_anio
rename P2K_MES p2k_mes
rename P2K p2k
rename P3A p3a
rename P3B p3b
rename P3C1 p3c1
rename P3C2 p3c2
rename P3C3 p3c3
rename P3C4 p3c4
rename P3C9 p3c9
rename P3D p3d
rename P3E p3e
rename P3F1 p3f1
rename P3F2 p3f2
rename P3G1_1 p3g1_1
rename P3G1_2 p3g1_2
rename P3G2_1 p3g2_1
rename P3G2_2 p3g2_2
rename P3G3_1 p3g3_1
rename P3G3_2 p3g3_2
rename P3G4_1 p3g4_1
rename P3G4_2 p3g4_2
rename P3G9 p3g9
rename P3G_TOT p3g_tot
rename P3H p3h
rename P3I p3i
rename P3J1 p3j1
rename P3J2 p3j2
rename P3K1 p3k1
rename P3K2 p3k2
rename P3K3 p3k3
rename P3K4 p3k4
rename P3K5 p3k5
rename P3K9 p3k9
rename P3L p3l
rename P4 p4
rename P4_1 p4_1
rename P4_2 p4_2
rename P4_3 p4_3
rename P4A p4a
rename P4A_1 p4a_1
rename P4B p4b

```

rename P4C p4c	rename P5D_TDIA p5d_tdia
rename P4D1 p4d1	rename P5E p5e
rename P4D2 p4d2	rename P5F1 p5f1
rename P4D3 p4d3	rename P5F2 p5f2
rename P4E p4e	rename P5F3 p5f3
rename P4F p4f	rename P5F4 p5f4
rename P4G p4g	rename P5F5 p5f5
rename P4H p4h	rename P5F6 p5f6
rename P4I p4i	rename P5F7 p5f7
rename P4I_1 p4i_1	rename P5F8 p5f8
rename P5 p5	rename P5F9 p5f9
rename P5A p5a	rename P5F10 p5f10
rename P5B_HLU p5b_hlu	rename P5F11 p5f11
rename P5B_MLU p5b_mlu	rename P5F12 p5f12
rename P5B_HMA p5b_hma	rename P5F13 p5f13
rename P5B_MMA p5b_mma	rename P5F14 p5f14
rename P5B_HMI p5b_hmi	rename P5F15 p5f15
rename P5B_MMI p5b_mmi	rename P5F99 p5f99
rename P5B_HJU p5b_hju	rename P6_1 p6_1
rename P5B_MJU p5b_mju	rename P6_2 p6_2
rename P5B_HVI p5b_hvi	rename P6_3 p6_3
rename P5B_MVI p5b_mvi	rename P6_4 p6_4
rename P5B_HSA p5b_hsa	rename P6_5 p6_5
rename P5B_MSA p5b_msa	rename P6_6 p6_6
rename P5B_HDO p5b_hdo	rename P6_7 p6_7
rename P5B_MDO p5b_mdo	rename P6_8 p6_8
rename P5B_THRS p5b_thrs	rename P6_9 p6_9
rename P5B_TDIA p5b_tdia	rename P6_10 p6_10
rename P5C p5c	rename P6_99 p6_99
rename P5D1 p5d1	rename P6A1 p6a1
rename P5D_HLU p5d_hlu	rename P6A2 p6a2
rename P5D_MLU p5d_mlu	rename P6A3 p6a3
rename P5D_HMA p5d_hma	rename P6A4 p6a4
rename P5D_MMA p5d_mma	rename P6A9 p6a9
rename P5D_HMI p5d_hmi	rename P6B1 p6b1
rename P5D_MMI p5d_mmi	rename P6B2 p6b2
rename P5D_HJU p5d_hju	rename P6C p6c
rename P5D_MJU p5d_mju	rename P6D p6d
rename P5D_HVI p5d_hvi	rename P7 p7
rename P5D_MVI p5d_mvi	rename P7A p7a
rename P5D_HSA p5d_hsa	rename P7B p7b
rename P5D_MSA p5d_msa	rename P7C p7c
rename P5D_HDO p5d_hdo	rename P8_1 p8_1
rename P5D_MDO p5d_mdo	rename P8_2 p8_2
rename P5D_THRS p5d_thrs	rename P8_3 p8_3
rename P8_4 p8_4	rename P9_1 p9_1
rename P8_9 p8_9	rename P9_H1 p9_h1
rename P8A p8a	rename P9_M1 p9_m1

```
rename P9_2 p9_2
rename P9_H2 p9_h2
rename P9_M2 p9_m2
rename P9_3 p9_3
rename P9_H3 p9_h3
rename P9_M3 p9_m3
rename P9_4 p9_4
rename P9_H4 p9_h4
```

```
rename P9_M4 p9_m4
rename P9_5 p9_5
rename P9_H5 p9_h5
rename P9_M5 p9_m5
rename P9_6 p9_6
rename P9_H6 p9_h6
rename P9_M6 p9_m6
rename P9_7 p9_7
rename P9_H7 p9_h7
rename P9_M7 p9_m7
rename P9_8 p9_8
rename P9_H8 p9_h8
rename P9_M8 p9_m8
```

\*guarde con sus correcciones pertinentes\*

\*¡Importante!

La base de datos ENOE correspondiente al año 2020, en el segundo trimestre no existe, por cuestiones sanitarias ese cuestionario se reformulo a "ETOE" (Encuesta Telefónica de Ocupación y Empleo), para fines de este trabajo, se ha omitido y la base de datos del año 2020 considera el trimestre I, III y IV, tome en cuenta que esto puede sesgar los datos. \*

\*Metodología append para unir trimestres y conformar base de datos anual\*

```
cd "E:\Usuarios\783251\Desktop\Steve Economia\Tesis\Datos\bases trimestrales"
```

```
us 2024_T1.dta, clear
```

```
/* drop cs_p20a_1 cs_p14_c cs_p21_des cs_p23_des
drop cs_p20a_c
drop cs_p20b_1
drop cs_p20b_c
drop cs_p20c_1 */
**opcional**
```

```
append using 2024_T2.dta
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\Sucias_años\2024_T12.dta"
```

```
append using 2024_T3.dta
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\Sucias_años\2024_T123.dta"
```

```
append using 2024_T4.dta
save "E:\Usuarios\783251\Desktop\Steve
Economia\Tesis\Datos\Sucias_años\2024.dta"
clear
```

```
***fin del código***
```

En este momento usted ha conformado la base de datos de la ENOE correspondiente al 2024, se sugiere que se guarden en una carpeta nueva donde puedas guardar únicamente las bases de datos anuales tome en consideración que el trabajo de tesis que se ha presentado es el armado de bases anuales de la ENOE de los años 2018, 2019, 2020, 2021, 2022, 2023 y 2024 usted puede hacer uso de esta base por años en estudios de cortes transversales por año e incluso general un panel de datos al integrar años a su base de datos, otra sugerencia resulta

que la base de datos resulta ser muy grandes por lo cual trabajar con esta puede ser muy tardado, algunas sugerencias tras la integración de esta base de datos es seleccionar mediante el comando keep las variables que estrictamente requiera, ya que la base de datos contempla cerca de 500 variables. Tome en cuenta los requerimientos técnicos, no olvide conservar variables importantes como AGEB o CVGEO para sus análisis espaciales y geográficos, así como los factores de expansión para el ajuste de modelos econométricos.

Las fichas técnicas y notas de ENOE se presentan en la información adjunta de los microdatos, por lo cual usted puede revisar a detalle la conformación y edición de las variables en los distintos años.

#### *Anexo B2. Fase preliminar; conformación de la muestra.*

Ya que la base de datos captura información indiscriminadamente, se debe seleccionar aquellas observaciones de interés, por ejemplo, aquellas que respondan el cuestionario, así como crear más variables como el ln del salario, así como las categorías de edades, grados escolares, la conformación, de la variable estado civil, la conformación de la población ocupada, entre otras, previamente seleccionadas y mencionadas en la metodología de esta tesis.

```
***Elección de variables***

tab r_def  /**Entrevista completa==0, revisar resultados
traslapados*/
  gen entrevista =0
  replace entrevista =1 if r_def==0
tab entrevista
drop if entrevista ==0

tab sex
tab eda

  rename eda edad
  gen grupo_edad=0
replace grupo_edad=6 if edad >=65
```

```

replace grupo_edad=5 if edad <=64
replace grupo_edad=4 if edad <=54
replace grupo_edad=3 if edad <=44
replace grupo_edad=2 if edad <=34
replace grupo_edad=1 if edad <=24
replace grupo_edad=0 if edad <=14
tab grupo_edad
label var grupo_edad "Grupo de edad"
label define lgrupo_edad 0 "Menor 14 años"
label define lgrupo_edad 1 "Menor 24 años",add
label define lgrupo_edad 2 "Menor 34 años",add
label define lgrupo_edad 3 "Menor 44 años",add
label define lgrupo_edad 4 "Menor 54 años",add
label define lgrupo_edad 5 "Menor 64 años",add
label define lgrupo_edad 6 "Mayor 65 años",add
label values grupo_edad lgrupo_edad
tab grupo_edad

```

```

gen edadmenor14 =0
replace edadmenor14 = 1 if grupo_edad==0
tab edadmenor14
gen edad14a24 =0
replace edad14a24 = 1 if grupo_edad==1
tab edad14a24
gen edad24a34 =0
replace edad24a34 = 1 if grupo_edad==2
tab edad24a34
gen edad34a44 =0
replace edad34a44= 1 if grupo_edad==3
tab edad34a44
gen edad44a54 =0
replace edad44a54 = 1 if grupo_edad==4
tab edad44a54
gen edad54a64 =0
replace edad54a64 = 1 if grupo_edad==5
tab edad54a64
gen edadmayor64 =0
replace edadmayor= 1 if grupo_edad==6
tab edadmayor

```

```

tab cs_p12
rename cs_p12 Sabe_leer_escribir
tab Sabe_leer_escribir

```

```

tab cs_p13_1
rename cs_p13_1 grado_esc
tab grado_esc

```

```

gen grado_escolar=0
replace grado_escolar= 1 if grado_esc == 0 | grado_esc == 1 ///
| grado_esc == 99 | grado_esc == .
replace grado_escolar= 2 if grado_esc == 2
replace grado_escolar= 3 if grado_esc == 3
replace grado_escolar= 4 if grado_esc == 4
replace grado_escolar= 5 if grado_esc == 5 | grado_esc == 6 ///
| grado_esc == 7 | grado_esc == 8 | grado_esc == 9

label var grado_escolar "Ultimo grado escolar finalizado"
label define lgrado_escolar 1 "Menor a Primaria", add
label define lgrado_escolar 2 "Primaria", add
label define lgrado_escolar 3 "Secundaria", add
label define lgrado_escolar 4 "Preparatoria", add
label define lgrado_escolar 5 "Profesionista", add
label values grado_escolar lgrado_escolar
tab grado_escolar

tab cs_p13_2
rename cs_p13_2 años_esc
tab años_esc

tab n_hij
gen n_hijos3=0
replace n_hijos3=1 if n_hij>=1 y n_hij<=3

label var n_hijos3 "Menos de 3 hijos"
label define ln_hijos3 0 "Tiene más de 4 hijos o No reportó", add
label define ln_hijos3 1 "Tiene de 1 a 3 hijos", add
label values n_hijos3 ln_hijos3
tab n_hijos3

gen n_hijos6=0
replace n_hijos6=1 if n_hij>=4 y n_hij<=26
label var n_hijos6 "Más de 3 hijos"
label define ln_hijos6 0 "Tiene menos de 3 hijos o No reportó",
add
label define ln_hijos6 1 "Tiene de 4 a 25 hijos", add
label values n_hijos6 ln_hijos6
tab n_hijos6

tab e_con /*Edo Conyugal*/
gen e_civil=0
replace e_civil=1 if e_con==1 | e_con==6
replace e_civil=0 if e_con==5 y e_con==2 y e_con==3 y e_con==4

```

```

label var e_civil "Estado Civil"
label define le_civil 0 "Soltero", add
label define le_civil 1 "Casado", add
label values e_civil le_civil
tab e_civil

tab zona
gen zonaA=0
replace zonaA=1 if zona ==1
tab zonaA
gen zonaB=0
replace zonaB=1 if zona ==2
tab zonaB

sum ingocup
bysort sex: sum ingocup

***Eliminamos Población***

/*list grupo_edad ingocup lsalario in 1/10
drop if ingocup==0 | edadmenor14==1*/

d ingocup
gen lsalario=ln(ingocup)

tab clase1

bysort sex: sum ingocup

tab clase1
tab clase2
tab clase3

*** segundo filtro ***

gen ocupado=0
replace ocupado=1 if clase3==1
tab ocupado
label var ocupado "Se encuentra ocupado y remunerado"
label define locupado 0 "No Ocupado", add
label define locupado 1 "Ocupado", add
label values ocupado locupado
tab ocupado
tab seg_soc
tab rama_est1
gen ActPrim=0
replace ActPrim=1 if rama_est1==1

```

```

tab ActPrim

gen ActSec=0
replace ActSec=1 if rama_est1==2
tab ActSec

gen ActTer=0
replace ActTer=1 if rama_est1==3
tab ActTer

tab ing7c
tab dur9c
tab dur_est

tab medica5c
gen serv_med=0
replace serv_med=1 if medica5c==2 | medica5c==3
replace serv_med=0 if medica5c==1 | medica5c==4 | medica5c==5
label var serv_med "Servicios Médicos"
label define lserv_med 0 "No tiene acceso", add
label define lserv_med 1 "Si tiene acceso", add
label values serv_med lserv_med
tab serv_med

tab sub_o
tab niv_ins

** ** ** **
drop if clase1 ==0
drop if rama_est2 ==0

*****Hasta aquí está contemplado el filtro de la PEA es decir de
la población que registro ingresos*****

// contemplan únicamente el sí registro. Eliminamos la Pob q no
aplica//

bysort sex: sum ingocup

*****
//tablas descriptivas (asigne filtros) *****

bysort sex: sum edad
bysort sex: tab e_civil
bysort sex: sum e_civil
bysort sex: tab ur
bysort sex: sum ur

```

```
bysort sex: tab Sabe_leer_escribir
bysort sex: sum Sabe_leer_escribir
```

```
bysort sex: tab grado_escolar
```

```
bysort sex: tab grupo_edad
```

```
*para el salario*
bysort sex: sum ingocup
```

```
///save "2023_limpia.dta"
```

```
**ya que se editó la base de datos se deberá guardar nuevamente con
las correcciones y ediciones pertinentes, a las cuales se les agrega
el sufijo "_limpias" esto hágase para cada una de las bases anuales
previas a los modelos**
```

```
***Fin del código***
```

### *Anexo C. Modelo Oaxaca Blinder, y Heckman*

```
*****Descomposición Oaxaca-Blinder*****
```

```
***factor expansión***
```

```
svyset [pw=est] /*pob total mayo a la real*/
```

```
*****
```

```
***Modelo Agregado***
```

```
oaxaca lsalario ur edadmayor edad54a64 edad44a54 edad34a44
edad24a34 edad14a24 edadmenor14 Sabe_leer_escribir grado_esc
e_civil seg_soc ActPrim ActSec ActTer ing7c serv_med niv_ins,
by(sex) noisily svy
```

```
* Var: lsalario ur grupo_edad Sabe_leer_escribir grado_esc
años_esc e_civil seg_soc rama ing7c dur_est serv_med niv_ins
anios_esc
```

```
***Problema num de hijos se recomienda crear var dicotomica
oaxaca, eform
```

```
*****
```

```
*****
```

```
oaxaca lsalario ur grupo_edad Sabe_leer_escribir grado_esc
años_esc e_civil seg_soc rama ing7c dur_est serv_med niv_ins
anios_esc, by(sex) pooled svy
```

```
oaxaca, eform
```

```
*****
```

```
tab clase3
```

```
*****
```

```
***heckman***
```

```
oaxaca lsalario ur grupo_edad grado_esc años_esc e_civil seg_soc
rama ing7c dur_est serv_med, by(sex) model2(heckman, twostep
select(ocupado= Sabe_leer_escribir niv_ins n_hijos3 n_hijos6))
```

```
by sex, sort : tabstat lsalario salario ur grupo_edad
Sabe_leer_escribir grado_esc años_esc e_civil seg_soc rama ing7c
dur_est serv_med niv_ins años_esc, s(n min max median mean sd sem
v cv p25 p75 p99)
```

```
*****reg agregada*****
```

```
by sex, sort : regress lsalario ur grupo_edad Sabe_leer_escribir
grado_esc años_esc e_civil seg_soc rama ing7c dur_est serv_med
niv_ins años_esc [pw=fac]
avplots
```

```
***Fin del código***
```

Mediante estos aportes se presentan las tablas de información correspondientes al capítulo 4 de los modelos Oaxaca en sus diferentes presentaciones, la metodología implementada se presenta previamente en el apartado de “metodología”.

Todo este código fue ejecutado en el software estadístico de Stata 17.

#### *Anexo D. Modelo SLX implementación de econometría espacial*

El modelo SLX se hizo bajo la implementación del Software estadístico R-studio, conformando el panel de datos de los años 2018-2024, exportando esta base generada al este paquete estadístico.

A continuación, se presenta el código para ejecutar este modelo, del cual se añaden los comentarios totales como la instalación de paqueterías y características técnicas, que se omiten a detalle su implementación usos generales y específicos, limitándose únicamente a la creación del modelo.

```
#Paso 1: Preparar las librerías
# Instalar si no lo tienes
```

```

install.packages("haven")

install.packages("dplyr")
install.packages("haven")      # para importar .dta
install.packages("DescTools") # para winsorizar
install.packages("tidyr")
install.packages("spdep")
install.packages("plm")
install.packages("ggplot2")

# Cargar librería
library(dplyr)
library(haven)      # para importar .dta
library(DescTools) # para winsorizar
library(tidyr)
library(spdep)
library(plm)
library(ggplot2)

#Paso 2: Importar la base
getwd()
setwd("E:/Usuarios/783251/Desktop/Steve
Economia/Tesis/Datos/Sucias_años")

df <- read_dta("E:/Usuarios/783251/Desktop/Steve
Economia/Tesis/Datos/Sucias_años/panel_municipal_oaxaca.dta")

# Paso 3: Winsorizar el ingreso y tomar logaritmo

p01 <- quantile(df$ingocup, 0.01, na.rm = TRUE)
p99 <- quantile(df$ingocup, 0.99, na.rm = TRUE)

df <- df %>%
  filter(ingocup > 0) %>%
  mutate(
    ing_winsor = ifelse(ingocup < p01, p01,
                       ifelse(ingocup > p99, p99, ingocup)),
    log_ing = log(ing_winsor)
  )
head(df)

#Paso 4: Calcular ingresos por sexo y municipio-año
#Crear tabla de ingresos para hombres
df_H <- df %>%
  filter(sex == 1, !is.na(log(ingocup + 1))) %>%
  group_by(cvegeog, cve_mun_num, year) %>%
  summarise(
    log_ing_H = mean(log(ingocup + 1), na.rm = TRUE),

```

```

    anios_esc_H = mean(anios_esc, na.rm = TRUE),
    edad_H = mean(eda, na.rm = TRUE),
    formal_H = mean(formal, na.rm = TRUE),
    seguridadsocial_H = mean(seguridadsocial, na.rm = TRUE),
    casado_union_H = mean(casado_union, na.rm = TRUE),
    urbano_H = mean(urbano, na.rm = TRUE),
    prestaciones_H = mean(prestaciones, na.rm = TRUE),
    accesosalud_H = mean(accesosalud, na.rm = TRUE),
    rama_industria_H = mean(rama_industria, na.rm = TRUE),
    rama_comercio_H = mean(rama_comercio, na.rm = TRUE),
    rama_servicios_H = mean(rama_servicios, na.rm = TRUE),
    rama_agro_H = mean(rama_agro, na.rm = TRUE),
    rama_constru_H = mean(rama_constru, na.rm = TRUE),
    rama_otro_menor_H = mean(rama_otro_menor, na.rm = TRUE),
    .groups = "drop"
  )

#Crear tabla de ingresos para mujeres
df_M <- df %>%
  filter(sex == 2, !is.na(log(ingocup + 1))) %>%
  group_by(cvegeo, cve_mun_num, year) %>%
  summarise(
    log_ing_M = mean(log(ingocup + 1), na.rm = TRUE),
    anios_esc_M = mean(anios_esc, na.rm = TRUE),
    edad_M = mean(eda, na.rm = TRUE),
    formal_M = mean(formal, na.rm = TRUE),
    seguridadsocial_M = mean(seguridadsocial, na.rm = TRUE),
    casado_union_M = mean(casado_union, na.rm = TRUE),
    urbano_M = mean(urbano, na.rm = TRUE),
    prestaciones_M = mean(prestaciones, na.rm = TRUE),
    accesosalud_M = mean(accesosalud, na.rm = TRUE),
    rama_industria_M = mean(rama_industria, na.rm = TRUE),
    rama_comercio_M = mean(rama_comercio, na.rm = TRUE),
    rama_servicios_M = mean(rama_servicios, na.rm = TRUE),
    rama_agro_M = mean(rama_agro, na.rm = TRUE),
    rama_constru_M = mean(rama_constru, na.rm = TRUE),
    rama_otro_menor_M = mean(rama_otro_menor, na.rm = TRUE),
    .groups = "drop"
  )

# Unir hombres y mujeres, y calcular la brecha
df_brecha <- inner_join(df_H, df_M, by = c("cvegeo",
" cve_mun_num", "year")) %>%
  mutate(
    brecha_log = log_ing_H - log_ing_M,
    esc_gap = anios_esc_H - anios_esc_M,
    edad_gap = edad_H - edad_M,

```

```

formal_gap = formal_H - formal_M,
seguridadesocial_gap = seguridadesocial_H - seguridadesocial_M,
urbano_gap = urbano_H - urbano_M,
prestaciones_gap = prestaciones_H - prestaciones_M,
accesosalud_gap = accesosalud_H - accesosalud_M,
rama_industria_gap = rama_industria_H - rama_industria_M,
rama_comercio_gap = rama_comercio_H - rama_comercio_M,
rama_servicios_gap = rama_servicios_H - rama_servicios_M,
rama_agro_gap = rama_agro_H - rama_agro_M,
rama_constru_gap = rama_constru_H - rama_constru_M,
rama_otro_menor_gap = rama_otro_menor_H - rama_otro_menor_M
)

# Vista preliminar
head(df_brecha)
saveRDS(df_brecha, "E:/Usuarios/783251/Desktop/Steve
Economia/Tesis/Datos/Sucias_años/df_brecha.rds")
df_brecha %>%
  summarise(across(ends_with("_gap"), ~ var(., na.rm = TRUE)))

#Gráfico de densidad de ingresos (logaritmo) por sexo
# Transformar a formato largo
df_ingresos <- df_brecha %>%
  select(cvegeo, year, log_ing_H, log_ing_M) %>%
  pivot_longer(cols = c(log_ing_H, log_ing_M),
               names_to = "sexo", values_to = "log_ing") %>%
  mutate(sexo = recode(sexo,
                       "log_ing_H" = "Hombres",
                       "log_ing_M" = "Mujeres"))

library(ggplot2)

# Gráfico con colores definidos
ggplot(df_ingresos, aes(x = log_ing, fill = sexo, color = sexo)) +
  geom_density(alpha = 0.3, size = 1.2) +
  scale_fill_manual(values = c("Hombres" = "#1f77b4", "Mujeres" =
"#e377c2")) +
  scale_color_manual(values = c("Hombres" = "#1f77b4", "Mujeres" =
"#e377c2")) +
  labs(
    title = "Distribución municipal del ingreso promedio (log) por
sexo",
    x = "Logaritmo del ingreso promedio",
    y = "Densidad",
    fill = "Sexo",
    color = "Sexo"
  ) +
  theme_minimal(base_size = 14) +

```

```

    theme(legend.position = "top")
# Crear tabla de varianzas ordenadas
tabla_varianzas_gap <- df_brecha %>%
  summarise(across(ends_with("_gap"), ~ var(., na.rm = TRUE))) %>%
  pivot_longer(everything(), names_to = "variable", values_to =
"varianza") %>%
  arrange(desc(varianza))

# Mostrar tabla
print(tabla_varianzas_gap)

#Paso 5. Leer shapefile y filtrar

install.packages("sf", dependencies = TRUE, type = "binary")

library(sf)
library(dplyr)

# Cargar shapefile de todos los municipios de México
shp <- st_read(
  dsn = "E:/Usuarios/783251/Desktop/Steve
Economia/Tesis/Datos/Sucias_años/series/",
  layer = "00mun"
) %>%
  st_transform(4326) %>%
  mutate(CVEGEO = as.integer(CVEGEO)) # identificador único por
municipio

carp <- "D:/Articulos/Oaxaca Blinder/2024/Sucias_años"
list.files(carp,
pattern="^inegi_refcenmuni_2010\\. (shp|dbf|shx|prj)$",
ignore.case=TRUE)
# Deben existir (al menos) .shp, .dbf y .shx juntos

"ESRI Shapefile" %in% sf::st_drivers()$name ## verificación de
archivos colaterales obli es TRUE

##Verificar correspondencia con tu base de datos
interseccion <- intersect(df_brecha$cvegeo, shp$CVEGEO)

cat("🌿 Municipios coincidentes:", length(interseccion), "\n")
cat("🇲🇽 Municipios en df_brecha:", n_distinct(df_brecha$cvegeo),
"\n")
cat("🏠 Municipios en shapefile:", n_distinct(shp$CVEGEO), "\n")

```

```

# Convertimos el identificador a integer en ambas bases por
seguridad
df_brecha <- df_brecha %>%
  mutate(cvegeo = as.integer(cvegeo))

shp <- shp %>%
  mutate(CVEGEO = as.integer(CVEGEO))

# Filtrar el shapefile para conservar solo municipios que están en
df_brecha
shp_filtrado <- shp %>%
  filter(CVEGEO %in% unique(df_brecha$cvegeo))

# Confirmar número de municipios comunes
cat("✅ Municipios comunes:", length(intersect(df_brecha$cvegeo,
shp_filtrado$CVEGEO)), "\n")

#Paso 6. Calcular los rezagos espaciales

# Calcular rezagos espaciales año por año
# 🧠 Función para calcular rezagos espaciales por año

# ✅ Función robusta para calcular rezagos espaciales por año

calcular_rezagos_por_año <- function(df, shp, variables, snap =
1e-5) {
  rezagos <- list()
  años <- sort(unique(df$year))

  for (yr in años) {
    message("📅 Procesando año: ", yr)

    df_t <- df %>%
      filter(year == yr, cvegeo %in% shp$CVEGEO) %>%
      arrange(cvegeo)

    shp_t <- shp %>%
      filter(CVEGEO %in% df_t$cvegeo) %>%
      arrange(CVEGEO)

    # 📧 Emparejamiento estricto por ID común
    comunes <- intersect(df_t$cvegeo, shp_t$CVEGEO)
  }
}

```

```

    df_t <- df_t %>% filter(cvegeo %in% comunes) %>%
arrange(cvegeo)
    shp_t <- shp_t %>% filter(CVEGEO %in% comunes) %>%
arrange(CVEGEO)

# ✅ Validación: deben coincidir perfectamente
if (!identical(df_t$cvegeo, shp_t$CVEGEO)) {
  warning("❌ ID de municipios desalineados en el año ", yr)
  next
}

# 🏠 Crear matriz de pesos espaciales
vecinos <- poly2nb(shp_t, snap = snap)
if (any(card(vecinos) == 0)) {
  warning("⚠️ Año ", yr, ": algunos municipios sin vecinos. Se
usará zero.policy = TRUE")
}

W_t <- nb2listw(vecinos, style = "W", zero.policy = TRUE)

# 📊 Calcular rezagos para cada variable
for (var in variables) {
  x <- df_t[[var]]
  x[!is.finite(x)] <- 0
  df_t[[paste0("WX_", var)]] <- lag.listw(W_t, x, zero.policy
= TRUE)
}

rezagos[[as.character(yr)]] <- df_t
}

bind_rows(rezagos)
}

#### Matriz de correlación solo para variables _gap
library(corrplot)

mat_corr_gap <- df_brecha %>%
  select(ends_with("_gap")) %>%
  cor(use = "pairwise.complete.obs")

# Visualizar
corrplot::corrplot(mat_corr_gap, method = "color", type = "upper",
tl.cex = 0.8)

# Extraer pares con |cor| > 0.5

```

```

mat_corr_gap[upper.tri(mat_corr_gap, diag =
FALSE)][abs(mat_corr_gap[upper.tri(mat_corr_gap)]) > 0.5]

library(reshape2)

cor_df <- melt(mat_corr_gap)
cor_df <- cor_df %>%
  filter(abs(value) > 0.5, Var1 != Var2) %>%
  arrange(desc(abs(value)))

print(cor_df)

vars_gap_final <- c(
  "esc_gap", "edad_gap", "formal_gap",
  "urbano_gap", # tiene varianza baja pero puede
ser importante territorialmente
  "rama_industria_gap", # si interesa estructura
productiva
  "rama_comercio_gap" # otra rama productiva
diferenciada
)

library(sf)
library(spdep)

df_full <- calcular_rezagos_por_año(df_brecha, shp_filtrado,
vars_gap_final)
saveRDS(df_full, "E:/Usuarios/783251/Desktop/Steve
Economia/Tesis/Datos/Sucias_años/df_full.rds")
saveRDS(df_full, "df_brecha_rezagos.rds")

names(df_full)
#####
##3
####visualización#####
##

df_2018_rez <- df_full %>% filter(year == 2018)
formal_2018 <- left_join(shp_filtrado, df_2018_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2018["WX_formal_gap"])

df_2019_rez <- df_full %>% filter(year == 2019)
formal_2019 <- left_join(shp_filtrado, df_2019_rez, by =
c("CVEGEO" = "cvegeo"))

```

```

plot(formal_2019["WX_formal_gap"])

df_2020_rez <- df_full %>% filter(year == 2020)
formal_2020 <- left_join(shp_filtrado, df_2020_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2020["WX_formal_gap"])

df_2021_rez <- df_full %>% filter(year == 2021)
formal_2021 <- left_join(shp_filtrado, df_2021_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2021["WX_formal_gap"])

df_2022_rez <- df_full %>% filter(year == 2022)
formal_2022 <- left_join(shp_filtrado, df_2022_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2022["WX_formal_gap"])

df_2023_rez <- df_full %>% filter(year == 2023)
formal_2023 <- left_join(shp_filtrado, df_2023_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2023["WX_formal_gap"])

df_2024_rez <- df_full %>% filter(year == 2024)
formal_2024 <- left_join(shp_filtrado, df_2024_rez, by =
c("CVEGEO" = "cvegeo"))
plot(formal_2024["WX_formal_gap"])

####
# Función para mapear un año
library(dplyr)
library(ggplot2)
library(sf)

# Agrupar todos los años en un solo data.frame
df_all_years <- df_full %>% filter(year %in% 2018:2024)

# Fusionar con el shapefile
mapa_all <- left_join(shp_filtrado, df_all_years, by = c("CVEGEO"
= "cvegeo"))

# Crear una variable categórica para la brecha
mapa_all <- mapa_all %>%
  mutate(
    brecha_cat = cut(
      WX_formal_gap,
      breaks = c(-Inf, -3, -2, -1, 0, 1, 2, 3, Inf),
      labels = c("< -3", "-3 to -2", "-2 to -1", "-1 to 0", "0 to
1", "1 to 2", "2 to 3", "> 3"),

```

```

        include.lowest = TRUE
      )
    )

# Crear el gráfico facetado
ggplot(mapa_all) +
  geom_sf(aes(fill = brecha_cat), color = NA) +
  facet_wrap(~year, ncol = 3) +
  scale_fill_manual(
    name = "Brecha formal",
    values = c(
      "< -3" = "#542788",
      "-3 to -2" = "#756bb1",
      "-2 to -1" = "#9e9ac8",
      "-1 to 0" = "#dadaeb",
      "0 to 1" = "#c7e9c0",
      "1 to 2" = "#74c476",
      "2 to 3" = "#238b45",
      "> 3" = "#00441b"
    ),
    na.value = "grey80"
  ) +
  labs(title = "Matriz de Pesos Espaciales de Empleo Formal 2018-
2024") +
  theme_minimal() +
  theme(
    legend.position = "right",
    strip.text = element_text(size = 12, face = "bold"),
    plot.title = element_text(hjust = 0.5, size = 16)
  )

# Municipios con mayor brecha espacial promedio (top 10)
top_municipios <- df_full %>%
  group_by(cvegeo) %>%
  summarise(brecha_espacial_prom = mean(WX_formal_gap, na.rm =
TRUE)) %>%
  arrange(desc(brecha_espacial_prom)) %>%
  slice_head(n = 10)

print(top_municipios)

write.csv(top_municipios, "Municipios_mayor_brecha.csv", row.names
= FALSE)

# Crear los plots
#p2018 <- plot_gap_year(2018)
#p2019 <- plot_gap_year(2019)

```

```

#p2020 <- plot_gap_year(2020)
#p2021 <- plot_gap_year(2021)
#p2022 <- plot_gap_year(2022)
#p2023 <- plot_gap_year(2023)
#p2024 <- plot_gap_year(2024)

# Ensamblar en dos láminas (2 filas, 4 columnas)
(p2018 | p2019 | p2020 | p2021) /
  (p2022 | p2023 | p2024 | plot_spacer())

#####
### Modelo SLX ###

vars_modelo_slx <- c(
  "brecha_log",
  "esc_gap", "edad_gap", "formal_gap", "urbano_gap",
  "rama_industria_gap", "rama_comercio_gap",
  "WX_esc_gap", "WX_edad_gap", "WX_formal_gap", "WX_urbano_gap",
  "WX_rama_industria_gap", "WX_rama_comercio_gap"
)

df_modelo <- df_full %>%
  select(cvegeo, year, all_of(vars_modelo_slx)) %>%
  drop_na()

# Convertir a panel
pdata <- pdata.frame(df_modelo, index = c("cvegeo", "year"))
saveRDS(pdata, "D:/Articulos/Oaxaca Blinder/2024/pdata.rds")
# Modelo SLX con efectos fijos
modelo_fx <- plm(
  brecha_log ~ esc_gap + edad_gap + formal_gap + urbano_gap +
    rama_industria_gap + rama_comercio_gap +
    WX_esc_gap + WX_edad_gap + WX_formal_gap + WX_urbano_gap +
    WX_rama_industria_gap + WX_rama_comercio_gap,
  data = pdata,
  model = "within",      # efectos fijos
  effect = "individual" # por municipio
)

summary(modelo_fx)

# Modelo SLX con efectos aleatorios
modelo_re <- plm(
  brecha_log ~ esc_gap + edad_gap + formal_gap + urbano_gap +
    rama_industria_gap + rama_comercio_gap +
    WX_esc_gap + WX_edad_gap + WX_formal_gap + WX_urbano_gap +
    WX_rama_industria_gap + WX_rama_comercio_gap,
  data = pdata,

```

```

    model = "random",
    effect = "individual"
)

summary(modelo_re)

##hausman_test <- phtest(modelo_fx, modelo_re)

hausman_test <- phtest(modelo_fx, modelo_re)
print(hausman_test)

###"Los resultados del modelo SLX con efectos fijos revelan que
las brechas de escolaridad (esc_gap)
#y formalidad (formal_gap) entre hombres y mujeres son factores
determinantes en la ampliación de
# la brecha salarial observada a nivel municipal. De manera
destacada, un incremento de un punto
# porcentual en la proporción de hombres empleados formalmente por
encima de las mujeres se asocia
#con un aumento del 55.8% en la brecha de ingresos, lo que refleja
una discriminación estructural en
#el acceso al empleo protegido. Asimismo, el rezago espacial de la
brecha de formalidad (WX_formal_gap)
# es estadísticamente significativo, sugiriendo la existencia de
mecanismos de difusión territorial de
#desigualdades de género."

library(caret)

nearZeroVar(df_modelo, saveMetrics = TRUE)

cor_matrix <- cor(df_modelo[, sapply(df_modelo, is.numeric)], use
= "pairwise.complete.obs")
round(cor_matrix, 2)

library(car)
modelo_lm_temp <- lm(
  brecha_log ~ .,
  data = df_modelo[, -which(names(df_modelo) %in% c("cvegeo",
"year"))]
)
vif(modelo_lm_temp)

##### Estimar el modelo SLX por año###

```

```

#Paso 1: Asegúrate de tener la matriz de pesos del año
correspondiente
library(spdep)
library(spatialreg)

# 1. Extraer municipios con datos en 2018
df_2018 <- df_full %>% filter(year == 2018) %>% arrange(cvegeo)
shp_2018 <- shp_filtrado %>% filter(CVEGEO %in% df_2018$cvegeo)
%>% arrange(CVEGEO)

# 2. Validar que los ID coincidan
stopifnot(identical(df_2018$cvegeo, shp_2018$CVEGEO))

# 3. Crear vecinos y matriz de pesos
vecinos_2018 <- poly2nb(shp_2018, snap = 1e-5)
W_2018 <- nb2listw(vecinos_2018, style = "W", zero.policy = TRUE)

#Paso 2: Estimar el modelo SLX para 2018
# 4. Especificar fórmula y ajustar el modelo
modelo_slx_2018 <- lmSLX(
  brecha_log ~ esc_gap + edad_gap + formal_gap + urbano_gap +
  rama_industria_gap + rama_comercio_gap,
  data = df_2018,
  listw = W_2018,
  zero.policy = TRUE
)

# 5. Resultados
summary(modelo_slx_2018)
r2 <- summary(modelo_slx_2018)$r.squared
r2
AIC(modelo_slx_2018)
BIC(modelo_slx_2018)
# Obtener residuos del modelo
residuos_slx <- residuals(modelo_slx_2018)

# Test de Moran
moran.test(residuos_slx, W_2018, zero.policy = TRUE)

#### Estimación SLX por año con evaluación del ajuste####
library(spdep)
library(spatialreg)
library(dplyr)

# Años del panel
anios <- 2018:2024

```

```

# DataFrame para almacenar resultados
resultados <- data.frame(
  Año = integer(),
  R2 = numeric(),
  AIC = numeric(),
  BIC = numeric(),
  Moran_I = numeric(),
  p_valor_Moran = numeric()
)

for (yr in anios) {
  cat("📊 Procesando año:", yr, "\n")

  df_anio <- df_full %>% filter(year == yr) %>% arrange(cvegeo)
  shp_anio <- shp_filtrado %>% filter(CVEGEO %in% df_anio$cvegeo)
  %>% arrange(CVEGEO)

  # Asegurar alineación
  stopifnot(identical(df_anio$cvegeo, shp_anio$CVEGEO))

  vecinos <- poly2nb(shp_anio, snap = 1e-5)
  W <- nb2listw(vecinos, style = "W", zero.policy = TRUE)

  # Estimar SLX
  modelo_slx <- lmSLX(
    brecha_log ~ esc_gap + edad_gap + formal_gap + urbano_gap +
      rama_industria_gap + rama_comercio_gap,
    data = df_anio,
    listw = W,
    zero.policy = TRUE
  )

  # Modelo OLS con WX para R2 aproximado
  modelo_ols <- lm(
    brecha_log ~ esc_gap + edad_gap + formal_gap + urbano_gap +
      rama_industria_gap + rama_comercio_gap +
      WX_esc_gap + WX_edad_gap + WX_formal_gap + WX_urbano_gap +
      WX_rama_industria_gap + WX_rama_comercio_gap,
    data = df_anio
  )

  r2 <- summary(modelo_ols)$r.squared
  aic <- AIC(modelo_slx)
  bic <- BIC(modelo_slx)

  residuos <- residuals(modelo_slx)
  moran <- moran.test(residuos, W, zero.policy = TRUE)
}

```

```

resultados <- rbind(resultados, data.frame(
  Año = yr,
  R2 = round(r2, 4),
  AIC = round(aic, 2),
  BIC = round(bic, 2),
  Moran_I = round(as.numeric(moran$estimate[1]), 4),
  p_valor_Moran = round(moran$p.value, 4)
))
}

# Mostrar tabla
print(resultados)

# Guardar como CSV
write.csv(resultados, "resultados_SLX_anuales.csv", row.names =
FALSE)

unique(df_brecha$cvegeo) %>% length()

library(tmap)
library(dplyr)

# Crear shapefile con indicador binario
shp_validacion <- shp %>%
  mutate(cvegeo = as.integer(CVEGEO)) %>%
  mutate(en_base = ifelse(cvegeo %in% unique(df_brecha$cvegeo), 1,
0))

# Activar vista interactiva si se desea
# tmap_mode("view") # <- para vista interactiva
tmap_mode("plot") # vista estática para publicación o revisión

# Mapa
tm_shape(shp_validacion) +
  tm_polygons("en_base",
    palette = c("lightgrey", "darkgreen"),
    labels = c("No disponible", "Municipio con datos"),
    title = "Presencia en base") +
  tm_layout(
    main.title = "Cobertura geográfica de la base df_brecha",
    main.title.size = 1.2,
    legend.outside = TRUE
  )
##Fin del código##

```