



# Brechas salariales por autoidentificación indígena y rasgos lingüísticos en México\*

*Wage gaps due to indigenous self-identification  
and linguistic traits in Mexico*

Eva O. Arceo-Gómez  
Universidad Iberoamericana, Departamento de Economía  
eva.arceo@ibero.mx

Pedro J. Torres L.  
Universidad Iberoamericana, Departamento de Economía

---

\*Agradecemos los comentarios de Gabriela Luna, Thibaut Plassot, Irving Rosales, Isidro Soloaga y dos dictaminadores anónimos. Los errores restantes son solo nuestros.

## Resumen

México es un país altamente desigual y el origen étnico-racial es una de las dimensiones que contribuyen a esta desigualdad. Estudios anteriores han analizado las diferencias salariales por origen-étnico racial y rasgos lingüísticos, pero hasta ahora ninguno ha reconocido que la autoidentificación indígena puede llevar a sesgos de selección en el análisis. En este artículo resolvemos este problema mediante la estimación de un modelo con cambio de régimen, el cual estima, en una primera etapa, la selección en la autoidentificación y, en una segunda, dos ecuaciones de los salarios según cómo se autoidentificó el trabajador. El modelo revela que la mayor parte del diferencial en el promedio de los salarios entre aquellos que se identifican como indígenas y aquellos que no se debe al sesgo de selección. A pesar de esto persiste un diferencial que no se explica por las características observadas ni la autoselección de estas subpoblaciones, el cual se podría deber a discriminación. Estos resultados contrastan con descomposiciones de Oaxaca tradicionales en que no se considera el sesgo de selección y apuntan a que la educación es el factor que más contribuye a la desigualdad salarial étnico-racial.

**Palabras clave:** brecha salarial; origen étnico; rasgos lingüísticos; México; regresión con cambio de régimen; autoidentificación indígena.

## Abstract

Mexico is a highly unequal country, and ethno-racial origin is one of the dimensions contributing to this inequality. Previous studies have analyzed wage differences by ethno-racial origin and linguistic traits. Still, so far, no study has recognized that indigenous self-identification can lead to selection biases in the analysis. In this article, we solve this problem by estimating a switching regression model, which in a first stage estimates the selection in self-identification, and in a second stage, two wage equations according to how the worker self-identified. The model reveals that most of the difference in average wages between those who identify as indigenous and those who do not is due to selection bias. Despite this, an unexplained differential persists (unexplained by observed characteristics or self-selection), which could be due to discrimination. These results contrast with traditional Oaxaca decompositions in which selection bias is not considered and suggest that education is the factor that contributes the most to ethnic-racial wage inequality.

**Keywords:** wage gap; ethnic origin; linguistic traits; Mexico; switching regression; indigenous self-identification.

**Clasificación JEL:** J15, J31, J71, O54.

**Fecha de recepción:** 11 de mayo de 2021

**Fecha de aceptación:** 28 de junio de 2021

## 1. Introducción

México es un país con altos niveles de desigualdad (El Colegio de México, 2018). A pesar de que 23% de la población se autoidentifica como indígena y que 7% habla una lengua indígena, la literatura en economía no ha explorado a profundidad el origen étnico como una causal de esta desigualdad.<sup>1</sup> Usando los datos de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) de 2018, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) estimó que casi 72% de la población indígena tenía ingresos por debajo de la línea de pobreza; por el contrario, solo 46% de la población no indígena vive con ingresos menores a la línea de pobreza (Coneval, 2021). Una fuente de esta desigualdad en los niveles de vida son las remuneraciones laborales, las cuales constituyen alrededor de 70% de los ingresos totales de los hogares (ENIGH, 2018).<sup>2</sup> En contraste con las brechas salariales por género, las brechas salariales por origen étnico de la población han sido poco exploradas en las ciencias económicas (Aguilar-Rodríguez et al., 2018; Canedo, 2019; Cano-Urbina y Mason, 2016; Hisamatsu y Ukeda, 2002, constituyen notables excepciones).

El objetivo de este artículo es estimar la brecha salarial por autoidentificación indígena, distinguiendo entre aquellos trabajadores que hablan una lengua indígena, los que solo hablan español y quienes hablan tanto una lengua indígena como español.<sup>3</sup> Como parte de nuestro análisis realizamos una descomposición de Oaxaca-Blinder (Blinder, 1973; Oaxaca, 1973) para estimar la proporción de la brecha que es explicada por diferencias en las características de la población autoidentificada como indígena y aquella no indígena. Sin embargo, la autoidentificación indígena puede estar ligada a características observables y no observables de los trabajadores, provocando así un sesgo de selección en las estimaciones de la brecha salarial. Para resolver esto, también estimamos un modelo

---

<sup>1</sup> De hecho, Advani et al. (2021) muestran que las disparidades raciales son mucho menos estudiadas en economía que en sociología o ciencias políticas; esto a pesar de que se publica mayor volumen en economía. Además, los economistas tienden a sobreestimar la cantidad de artículos publicados sobre cuestiones raciales en economía mucho más que en otras ciencias sociales.

<sup>2</sup> Esta cifra está basada en el promedio del ingreso total corriente y en el promedio del ingreso laboral de los hogares mexicanos.

<sup>3</sup> La autoidentificación o autoadscripción étnica se determinan a partir de la pregunta 13, Sección III, del cuestionario de hogares de la ENIGH 2018: “De acuerdo con la cultura de (NOMBRE), ¿ella (él) se considera indígena?”

de regresión con cambio de régimen (*switching regression model*), el cual modela, en su primera etapa, la selección de identificación indígena y, en su segunda etapa, estima dos modelos de la ecuación salarial correspondientes a la población indígena de trabajadores y a la población no indígena de trabajadores. La aplicación de este modelo de cambio de régimen para el caso de las brechas salariales por origen étnico, así como la descomposición de Oaxaca que se desprende de él, constituyen nuestra principal contribución a la literatura.

A raíz del imperialismo europeo en el siglo xv y del impacto de los procesos de colonización sobre las relaciones étnicas de las colonias, muchos países sufren disparidades étnico-raciales con respecto a los estándares de vida. En México, los pueblos originarios se encuentran en las zonas más marginadas, donde la constante son menores oportunidades educativas, laborales y, en general, de desarrollo. Hoy en día existen más de 60 culturas indígenas en México, alrededor de 19.14% de la población se autoidentifica como indígena y 6.1% habla una lengua nativa.<sup>4</sup> De acuerdo con el informe de “Diversidad étnico-racial en México y su influencia en la movilidad social” desarrollado por el Banco Interamericano de Desarrollo (Rojas Lomelín et al., 2019), la mayor concentración de las personas que se autoidentifican como indígenas se encuentra en la región sur del país, donde en su momento predominaron las culturas mesoamericanas y hubo una menor cantidad de asentamientos europeos. En los estados de Guerrero, Puebla, Hidalgo, Campeche, Chiapas, Quintana Roo, Oaxaca y Yucatán se reporta una proporción de población indígena mayor a 20% (Figura 1).

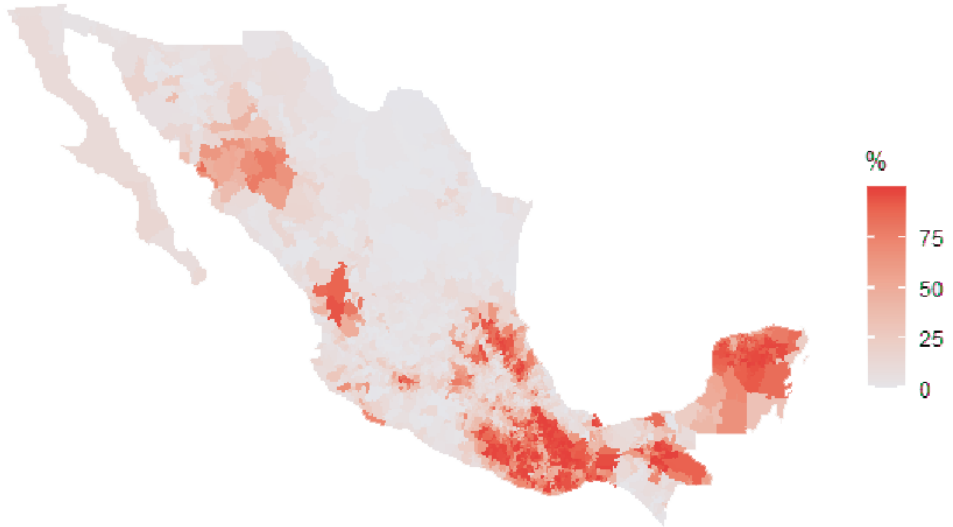
Diversos informes y estudios han encontrado que en México existe un rezago asociado a las características étnico-raciales de los individuos en materia de educación, ingresos y riqueza. En el tema educativo, factores como el tono de piel, el habla de una lengua indígena (ya sea por uno mismo o por los padres) y la autoadscripción étnica presentan tasas de acceso, rendimiento y terminación escolar inferiores (Suárez Blanch, 2006; Rojas Lomelín et al., 2019; Solís et al., 2019a). En particular, Solís et al. (2019b) encuentran que 43% de las personas hablantes de una lengua indígena no tiene primaria completa y solo 9% alcanza el nivel medio superior.

Además, la autoidentificación indígena y el habla de una lengua indígena se asocian a menores ingresos laborales, los cuales constituyen en promedio 70% de los ingresos corrientes totales de los hogares (ENIGH, 2018). De la población indígena que se encuentra ocupada, 35% no recibe un salario; en contraste, para

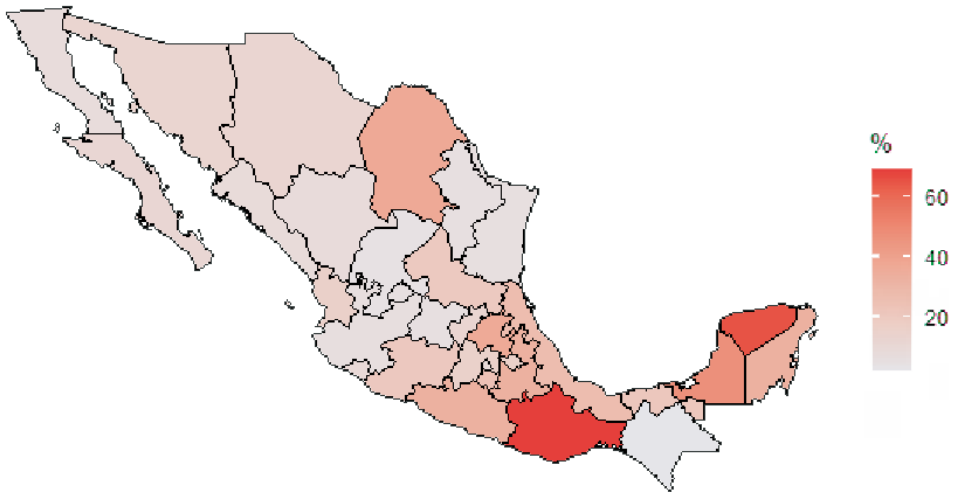
<sup>4</sup> Datos estimados del Censo de Población y Vivienda 2020.

**Figura 1. Porcentaje de la población que se identifica como indígena**

A. A nivel municipal



B. A nivel estatal



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2020 (INEGI, 2020).

Panel (A) muestra el porcentaje de la población que se identifica como indígena por municipio; mientras que el Panel (B) lo hace por estado.

la población total esta cifra se encuentra en 25%. Cano-Urbina y Mason (2016) estimaron que las personas que se consideran indígenas ganan en promedio 9% menos que la población no indígena, incluso después de controlar por años de educación, edad, ocupación e industria. Entre aquellos que se identifican como indígenas, los bilingües (español y una lengua indígena) perciben un mayor salario que los monolingües (solo lengua indígena); esto sería de esperarse dado que el idioma predominante en México es el español. Sin embargo, aquellos que solo hablan español perciben un mayor salario que quienes son bilingües (Gutiérrez Chong, 2012), lo que parece ser contra intuitivo con respecto a la teoría del capital humano: se esperaría que la lengua indígena, sin tener una ventaja para el mercado, cuando menos no perjudicaría.

Canedo (2019) estima que quienes no hablan lenguas indígenas reciben mayores retornos por su capital humano que los que sí hablan dichas lenguas, pero no incluye en su análisis a la autoidentificación indígena. También encuentra que esta desigualdad es mayor en los deciles bajos de la distribución de ingresos en el país. Canedo (2019) estima la brecha étnica en salarios promedio en alrededor de 82%. A través de una descomposición de Oaxaca-Blinder encuentra que solo 32% de esta brecha es explicada por diferencias en las características de los trabajadores, por lo que la autora infiere que gran parte de la brecha es debido a discriminación. Por otro lado, Hisamatsu y Ukeda (2002) indican que la discriminación salarial hacia los pueblos indígenas se explica en 85% por factores relacionados al ingreso, mientras que el otro 15% se debe a desventajas por la identificación cultural o características no observables como la discriminación. Estos resultados coinciden con lo encontrado por Aguilar-Rodríguez et al. (2018) y Horbath (2008).

Si bien todos los estudios enfocados en brechas salariales por origen étnico son incapaces de atribuir las diferencias únicamente a la discriminación, existen tres investigaciones que refieren el impacto del tono de piel en resultados de vida. En un mercado laboral muy particular, Campos-Vázquez (2021) midió el precio del tono de piel: escorts con tonos de piel más oscuros cobran 10% menos que aquellas con todos de piel más claros, después de controlar por la edad y otras características físicas. En un experimento controlado de campo, Arceo-Gómez y Campos-Vázquez (2014) encontraron que las mujeres con tonos de piel más oscuros reciben menos entrevistas de trabajo que quienes tienen tonos de piel clara. Finalmente, en un experimento de laboratorio, Campos-Vázquez et al. (2017) hallaron que los estereotipos asociados a la identidad indígena y al

tono de piel impactan, incluso, las aspiraciones de los adolescentes, lo cual, a su vez, podría incidir de manera negativa en sus resultados de vida. Los hallazgos de este experimento nos llevaron a cuestionarnos si la propia autoidentificación indígena podría sesgar los resultados de las estimaciones de brechas salariales. Además, la existencia de estas diferencias en investigaciones experimentales justifica el estudio de las brechas salariales, aunque estas no se puedan atribuir solo a la discriminación. Las brechas en sí son multicausales.

En este trabajo buscamos identificar las diferencias en los pagos relacionados con los diferentes aspectos del capital humano y rasgos lingüísticos en el salario de los individuos de acuerdo con su identificación indígena. Esto lo llevamos a cabo en tres pasos usando datos de ENIGH 2018. Primero, estimamos un modelo de mínimos cuadrados ordinarios (MCO), en el cual tomamos en cuenta identificación indígena y rasgos lingüísticos. A partir de este modelo estimamos una descomposición a la Oaxaca-Blinder de la brecha salarial, la cual difiere de Canedo (2019) por la inclusión de rasgos lingüísticos en la población. La descomposición de Oaxaca revela que todavía 60% de la brecha salarial se debe a diferencias en características, y casi todo el resto a diferencias en el retorno a dichas características.

En tanto que Cano-Urbina y Mason (2016) aprecian que las personas que dejan de identificarse como indígenas mejoran sus condiciones salariales en México, y relacionan este resultado a un proceso de aculturación, es necesario corregir por selección a la identidad indígena. El proceso de aculturación y de creación de una identidad podría ser estratégico y una decisión consciente de las personas. Ante la existencia de discriminación, las personas indígenas pueden adaptar su identidad a una norma social donde las no indígenas se consideran mejores, más preparadas, o más productivas. La propia existencia de tratamientos para aclarar la piel ilustra cómo las personas manipulan su tono de piel para adaptarse a una norma de belleza más aceptada. Desde un punto de vista económico, los individuos no incurrirían en este costo si no tuviese un retorno, el cual se puede dar en el mercado laboral, en el matrimonio u otros contextos sociales. Además, como hallaron Campos-Vázquez et al. (2017), la identidad indígena podría impactar las aspiraciones de las personas. Por ello, es importante tomar en cuenta que la autoidentificación indígena podría ser una variable endógena que induce sesgos de selección en las estimaciones que hallamos actualmente en la literatura. Aquí proponemos corregir posibles sesgos de selección a través de un modelo de regresión con cambio de régimen. En este, en una primera etapa, se corrige por la autoselección de los individuos de identificarse como indígenas, para así, en una

segunda, estimar las diferencias en los pagos a través de regresiones específicas para población indígena y no indígena. Adicionalmente, y en contraste con Canedo (2019), estimamos una descomposición tipo Oaxaca-Blinder que considera tanto la autoidentificación indígena como los rasgos lingüísticos de los trabajadores. Los resultados sugieren una brecha salarial con respecto a la identificación indígena de 8%; asimismo, encuentra que aquellos individuos que no se identifican como indígena, pero hablan una lengua indígena (es decir, que son bilingües), reciben mayores remuneraciones que aquellos que son bilingües y se identifican como indígenas. Con respecto a la educación, no se encuentra una diferencia a las remuneraciones del factor capital humano por identificación indígena. La corrección por selección provoca que las características tengan mayor peso en la descomposición de Oaxaca que en la descomposición con el modelo lineal.

En el apartado 2 presentamos los datos y un breve análisis descriptivo de las diferencias salariales entre la población indígena y no indígena. En el 3 abordamos la forma en la que se estima la brecha y la diferencia en los pagos. En el 4 mostramos los resultados de nuestras estimaciones de los modelos y en el 5 presentamos la robustez de nuestros resultados a distintos grupos de variables de control. Finalmente, en el 6 concluimos.

## 2. Datos y estadísticas descriptivas

Para determinar la brecha salarial en el mercado laboral mexicano se utilizaron datos de ENIGH 2018 colectados por el INEGI, representativos en el nivel nacional, por ámbito rural y urbano, y por entidad federativa. En esta encuesta se encuentran dos variables clave para el análisis propuesto: primero, tenemos información sobre la capacidad individual de hablar una lengua indígena (o español, preguntas 9 y 11, Sección III del Cuestionario de Hogares y Viviendas); segundo, nos proporciona información sobre la autoidentificación étnica de los individuos (pregunta 13, Sección III del Cuestionario de Hogares y Viviendas). Para fines de este estudio solo se diferencia entre aquellas personas que se identifican con algún pueblo indígena (indígenas) y aquellos que no (no indígenas). Únicamente se tomaron en cuenta los individuos que no presentan datos faltantes en las variables de interés. De igual manera, solo es de interés aquel grupo de personas que reportaron alguna remuneración laboral (sueldos, salarios o jornal; destajo; comisiones y propinas; horas extra). Por ello, nuestros resultados deben ser interpretados



de forma condicional a recibir una remuneración laboral. De la encuesta se tomaron los individuos entre veinticinco y sesenta y cinco años. La muestra final consta de 68 585 observaciones al nivel individuo. Dentro de estas personas, 38% son mujeres y la edad promedio es de cuarenta años. En la muestra, 29% se identifica como indígena, y 6% habla una lengua indígena, lo cual es cercano a los resultados de la Encuesta Intercensal 2015 (23% y 7%, respectivamente).

La Tabla 1 presenta la estadística descriptiva de algunas variables clave en el análisis para toda la población, así como por adscripción indígena. Las últimas dos columnas muestran la diferencia de medias entre no indígenas e indígenas y el valor *p* de la prueba de diferencia de medias. Podemos observar que únicamente la variable secundaria no presenta una diferencia estadísticamente significativa al 5% entre aquellos individuos que se identifican como indígenas y los que no, en la muestra de estimación. La diferencia de medias en la edad, aunque estadísticamente significativa, no es económicamente relevante. El resto de las características de los individuos, que se refieren a salarios, habla de lengua indígena y escolaridad, sí presentan diferencias relevantes y significativas. En particular, mostramos que quienes no se consideran indígenas ganan alrededor de 23% más por hora. Finalmente, existen personas que, aunque no se identifican como indígenas, hablan una lengua nativa y representan alrededor de 0.4% de la muestra.

**Tabla 1. Estadística descriptiva de la muestra de estimación**

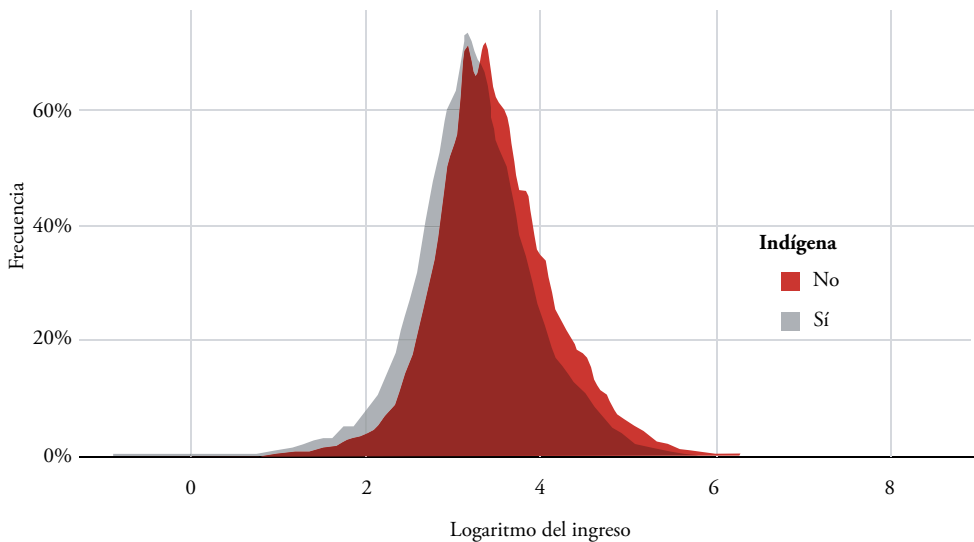
| <i>Variables</i>    | <b>Total</b> |                   | <b>Indígena</b> |                   | <b>No Indígena</b> |                   | <b>Diferencia</b> |                |
|---------------------|--------------|-------------------|-----------------|-------------------|--------------------|-------------------|-------------------|----------------|
|                     | <i>Media</i> | <i>Desv. Est.</i> | <i>Media</i>    | <i>Desv. Est.</i> | <i>Media</i>       | <i>Desv. Est.</i> |                   | <i>P-value</i> |
| <b>Salario</b>      | 39.78        | 53.69             | 33.44           | 52.49             | 42.39              | 53.95             | 8.95              | 0              |
| <b>Log(salario)</b> | 3.41         | 0.7               | 3.25            | 0.69              | 3.48               | 0.69              | 0.23              | 0              |
| <b>Indígena</b>     | 0.29         | 0.45              |                 |                   |                    |                   |                   |                |
| <b>Lengua</b>       | 0.06         | 0.23              | 0.18            | 0.38              | 0.004              | 0.06              | 0.176             | 0              |
| <b>Bilingüe</b>     | 0.05         | 0.22              | 0.17            | 0.38              | 0.004              | 0.06              | 0.166             | 0              |
| <b>Primaria</b>     | 0.29         | 0.45              | 0.35            | 0.48              | 0.26               | 0.44              | 0.09              | 0              |
| <b>Secundaria</b>   | 0.33         | 0.47              | 0.34            | 0.47              | 0.33               | 0.47              | 0.01              | 0.09           |
| <b>Preparatoria</b> | 0.21         | 0.41              | 0.18            | 0.39              | 0.22               | 0.41              | 0.04              | 0              |
| <b>Superior</b>     | 0.17         | 0.38              | 0.13            | 0.33              | 0.19               | 0.39              | 0.06              | 0              |
| <b>Edad</b>         | 40.47        | 10.41             | 40.32           | 10.46             | 40.52              | 10.39             | 0.20              | 0.02           |
| <b>Mujer</b>        | 0.38         | 0.49              | 0.37            | 0.48              | 0.38               | 0.49              | 0.01              | 0.04           |

Fuente: Elaboración propia con datos de ENIGH 2018.

La Tabla presenta estadísticas descriptivas básicas para las variables de interés.

La estadística descriptiva básica de las variables clave, en conjunto con la distribución porcentual del ingreso (Figura 2), nos da indicios sobre una posible brecha salarial con respecto a la identificación indígena. Asimismo, es posible observar que los individuos que no se consideran indígenas cuentan con mayores niveles de educación media superior y superior, lo cual, en teoría, presupone mayores remuneraciones en términos de salario. La prueba de Kolmogorov-Smirnov indica que existe una diferencia en la distribución del ingreso en ambos grupos. En específico, indica que la distribución del grupo indígena se encuentra a la izquierda de la distribución del grupo no indígena, lo cual es evidencia de que en la población indígena los ingresos laborales bajos son más probables que entre la no indígena.

Figura 2. Distribución del ingreso



Fuente: Elaboración propia con datos de la ENIGH 2018.

La Figura presenta la distribución del logaritmo del ingreso de los individuos según su identificación indígena. En gris se muestran las personas que se identifican como indígenas y en rojo las que no.

### 3. Metodología empírica

Nuestra metodología empírica consta de tres partes. Primero, estimamos mediante mínimos cuadrados ordinarios (MCO) la ecuación minceriana de los salarios, añadiendo la autoidentificación indígena, el habla de lengua indígena o el ser bilingüe a la regresión para calcular la brecha salarial promedio de este grupo. A partir de esta función lineal estimamos una descomposición Oaxaca-Blinder (Oaxaca, 1973), la cual evalúa qué sucedería si el grupo indígena tuviera las mismas características de capital humano que el no indígena. Esto nos permite descomponer las diferencias en tres grupos: *i*) diferencias por las características; *ii*) diferencias en los pagos (coeficientes), y *iii*) la diferencia por la interacción entre las características y los coeficientes. Al analizar estas tres diferencias es posible ver qué parte de la brecha es explicada por las características individuales y cuál puede ser atribuida a discriminación laboral. Subsecuentemente, estimamos un modelo de regresión con cambio de régimen (*switching regression*) para corregir por la autoidentificación indígena. Con este modelo se busca establecer la diferencia que hay en el pago a los individuos que se consideran indígenas y aquellos que no, por los diferentes niveles educativos y las lenguas que hablan. El modelo ajusta las estimaciones de las ecuaciones mincerianas de acuerdo con un modelo tipo Heckman, en el que se controla por selección muestral imputando la razón inversa de Mills (RIM). La RIM es resultado de un modelo Probit en una primera etapa, que estima la probabilidad de que un individuo se identifique como indígena.

Tomando como punto de partida los modelos empleados por Aguilar-Rodríguez et al. (2018), Canedo (2019), Cano-Urbina y Mason (2016), y Hisamatsu y Ukeda (2002), proponemos el siguiente modelo para medir la brecha salarial:

$$lw_i = \alpha_0 + \alpha_1 Indígena_i + \alpha_2 Lengua_i + \alpha_3 Bilingüe_i + \alpha_4 Indígena \times Bilingüe + \mathbf{X}_i \beta + \mu_i \quad (1)$$

donde  $lw_i$  es el logaritmo del salario del individuo ( $i$ ),  $Indígena_i$ , una variable dummy que indica si el individuo se identifica como indígena (1=si, 0=no);  $Lengua_i$ , una variable dummy que indica si el individuo habla una lengua indígena (1=si, 0=no);  $Bilingüe_i$ , una variable dummy que indica si el individuo habla una lengua indígena y español (1=si, 0=no);  $\mathbf{x}_i$ , un vector de características que determinan el salario,<sup>5</sup>

<sup>5</sup> Incluye las variables edad y edad,<sup>2</sup> variable dicotómica de mujer, variables dicotómicas del tamaño de localidad, variables dicotómicas del sector de actividad económica, y variables dicotómicas de ocupación.

y  $\mu_i$ , el término de error que se supone independiente e idénticamente distribuido con media cero y varianza constante.

Si no existe una brecha salarial con respecto a la identificación indígena, se espera que el coeficiente  $\alpha_1$  no sea estadísticamente significativo. Siguiendo la teoría del capital humano, se espera que el coeficiente relativo a ser bilingüe ( $\alpha_3$ ) tenga signo positivo (o al menos sea no negativo). Dado que la lengua predominante en México es el español, es de esperarse que, una vez controlado por ser bilingüe, el habla de una lengua indígena no sea estadísticamente significativo. Esto implicaría que el habla de una lengua indígena solo tiene un efecto sobre el salario si el individuo habla español.

Siguiendo a Oaxaca (1973), podemos observar el efecto que tiene la identificación indígena sobre el resto de las variables al estimar la regresión en dos grupos diferentes:

$$\begin{aligned} Y_{iN} &= \alpha_N + \mathbf{X}_{iN} \beta_N \\ Y_{iI} &= \alpha_I + \mathbf{X}_{iI} \beta_I \end{aligned} \quad (2)$$

donde el superíndice  $N$  denota la ecuación de los individuos que no se consideran indígenas y el superíndice  $I$  la ecuación de aquellos que sí. El promedio en el ingreso de cada grupo está dado por:

$$\bar{Y}_j = \alpha_j + \bar{\mathbf{X}}_j \beta_j, \forall j \in \{N, I\}.$$

Una vez divididos los grupos es posible estimar la diferencia en la media de ambos grupos por medio de la siguiente ecuación:

$$\Delta \bar{Y} = \bar{Y}_N - \bar{Y}_I = \widehat{\alpha}_N + \bar{\mathbf{X}}_N \widehat{\beta}_N - \widehat{\alpha}_I - \bar{\mathbf{X}}_I \widehat{\beta}_I$$

Esta expresión se puede reescribir como:

$$\Delta \bar{Y} = \underbrace{\widehat{\alpha}_N - \widehat{\alpha}_I}_U + \underbrace{(\bar{\mathbf{X}}_N - \bar{\mathbf{X}}_I) \widehat{\beta}_N}_E + \underbrace{\bar{\mathbf{X}}_I (\widehat{\beta}_N - \widehat{\beta}_I)}_C + \underbrace{(\bar{\mathbf{X}}_N - \bar{\mathbf{X}}_I) (\widehat{\beta}_N - \widehat{\beta}_I)}_{CE} \quad (3)$$

donde la diferencia total es  $E + C + CE + U$ ;  $E$  es la diferencia asociada a las características o diferencia explicada;  $C$ , la diferencia asociada a los coeficientes o retornos a las características;  $CE$ , la interacción entre ambas diferencias;  $U$ , la diferencia por factores no observables, y por tanto  $C + U$  sería la diferencia

atribuible a factores no explicados como la discriminación laboral, habilidades suaves y otras variables no observables en los datos.

En México la atribución indígena está definida por la autoidentificación más allá de los usos y costumbres, la región o el habla de una lengua indígena. La autoidentificación indígena podría generar un sesgo de selección en la estimación por MCO. Considerando esto, y extrapolando los modelos propuestos por Adamchik y Bedi (2000), van der Gaag y Vijverberg (1988), Lee (1978), y Robinson y Tomes (1984), un individuo decide si identificarse o no como indígena tomando en cuenta

$$\frac{U(w_i, Z_i)_{I_i} - U(w_i, Z_i)_{N_i}}{U(w_i, Z_i)_{N_i}} > \rho_i \approx \ln(U(w_i, Z_i)_{I_i}) - \ln(U(w_i, Z_i)_{N_i}) > \rho_i \quad (4)$$

si la diferencia en la utilidad, donde  $\ln(U(w_i, Z_i)_{I_i})$  es la utilidad de un individuo que se considera indígena y  $\ln(U(w_i, Z_i)_{N_i})$  es la utilidad del mismo individuo si no se identificara como indígena, es mayor a su utilidad de reserva ( $\rho_i$ ), el individuo decide identificarse como indígena. Matemáticamente, la utilidad de reserva se aproxima a:

$$\rho_i = \mathbf{X}_i \beta + \xi_i$$

sustituyendo y reordenando los términos obtenemos la siguiente expresión:

$$I_i^* = \left[ \ln(U(w_i, Z_i)_{I_i}) - \ln(U(w_i, Z_i)_{N_i}) \right] - [\mathbf{X}_i \beta + \xi_i] \quad (5)$$

donde  $I_i^*$  representa una variable latente que cuando  $I_i^* > 0$  el individuo se considera indígena y si  $I_i^* \leq 0$  no se identifica como indígena. Esta probabilidad se puede estimar en primera instancia por medio de un modelo Probit (que representa la forma reducida):<sup>6</sup>

$$\Pr(\text{Indígena}_i = 1 | Z_i) = \Phi(Z_i + v_i) \quad (6)$$

donde  $\Phi(\bullet)$  es un vector de características que influyen sobre la autoidentificación indígena,<sup>7</sup> es la función de distribución normal acumulada, y  $v_i$  es el término de error que se supone independiente e idénticamente distribuido con media

<sup>6</sup> Para mayor entendimiento sobre el modelo que estimamos ver Wooldridge (2010).

<sup>7</sup> Se puede observar que en este modelo la variable *Lengua* no está incluida, ya que servirá como característica para identificar la autoidentificación indígena.

cero y varianza constante. En la implementación de este modelo utilizaremos hablar una lengua indígena como la variable de exclusión en el modelo para la autoidentificación indígena.

Con estos resultados se imputa un factor de corrección de selección (la razón del inverso de Mills,  $RIM$ ) a la ecuación de salarios a estimar, resultando en:

$$\begin{aligned} lw_{N_i} &= X_i\gamma_i + \beta_{N_i}RIM_i + \xi_{N_i} \\ lw_{I_i} &= X_i\delta_i + \beta_{I_i}RIM_i + \xi_{I_i}, \end{aligned} \quad (7)$$

donde  $\gamma_i - \delta_i$  representa la diferencia en los pagos por la característica  $j$  que determina el salario, dependiendo de si un individuo se considera indígena o no.

#### 4. Resultados principales

La Tabla 2 presenta los resultados de la estimación de la ecuación minceriana por MCO (columna 1) y el cambio porcentual exacto dados los coeficientes (columna 2).<sup>8</sup> Podemos observar que hay diferencias significativas de ingreso por rasgos etnolingüísticos. La categoría base en nuestra estimación son aquellos que no se identifican como indígenas que solo hablan español. Así, nuestros resultados sugieren que, con respecto a esta categoría base, un no indígena bilingüe recibe un salario 30% mayor en promedio; un indígena que solo habla español, 4.2% menor; un indígena que solo habla su lengua nativa, 40% menor, y finalmente un indígena bilingüe recibe un salario 15% menor. Entonces, tanto la identificación indígena como el habla de una lengua indígena se relacionan con menores salarios de manera independiente. Sin embargo, es sorprendente que ser bilingüe suponga un mayor pago para quienes no se consideran indígenas. Solo alrededor de 200 observaciones en la muestra pertenecen a este último subgrupo, por lo cual podrían estar muy seleccionados del resto de la población no indígena.

Una vez analizada la relación entre los salarios y la identificación indígena se estudian los posibles canales por los que se genera esta brecha: características o retornos y no observables en el modelo. Los resultados de la descomposición Oaxaca-Blinder, junto con intervalos de confianza a 95% (estimados para 100 réplicas; Figura 3) estima la brecha salarial respecto a la identificación indígena en 20%, de la cual aproximadamente 6 puntos porcentuales (pp) provienen

<sup>8</sup> El cambio porcentual exacto se obtiene a partir de la siguiente expresión:  
 $(\exp^{\beta_i} - 1) * 100.$

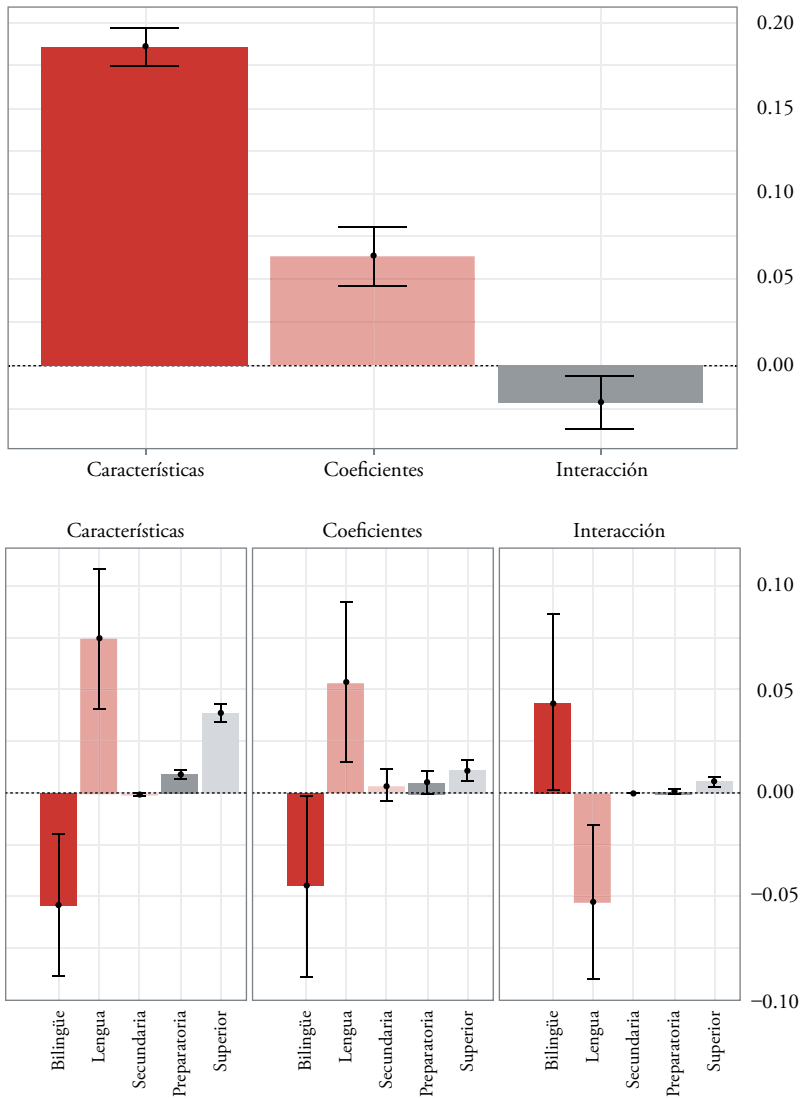
**Tabla 2: Resultados estimados por MCO**

|                             | Variable dependiente: Log(Salario) |                                 |
|-----------------------------|------------------------------------|---------------------------------|
|                             | Resultados<br>(1)                  | Cambio porcentual exacto<br>(2) |
| Constante                   | 2.887*** (0.038)                   |                                 |
| Indígena                    | -0.042*** (0.005)                  | -4.11                           |
| Lengua                      | -0.358*** (0.076)                  | -30.09                          |
| Bilingüe                    | 0.299*** (0.086)                   | 34.85                           |
| Indígena*Bilingüe           | -0.047 (0.041)                     | -4.59                           |
| <i>Controles:</i>           |                                    |                                 |
| <i>Edad</i>                 | Sí                                 |                                 |
| <i>Mujer</i>                | Sí                                 |                                 |
| <i>Entidad federativa</i>   | Sí                                 |                                 |
| <i>Tamaño de localidad</i>  | Sí                                 |                                 |
| <i>Nivel de escolaridad</i> | Sí                                 |                                 |
| <i>Sector</i>               | Sí                                 |                                 |
| <i>Ocupación</i>            | Sí                                 |                                 |
| Observaciones               | 68,585                             |                                 |
| R <sup>2</sup>              | 0.393                              |                                 |
| R <sup>2</sup> Ajustada     | 0.392                              |                                 |
| Error. Est.                 | 0.548 (df = 68513)                 |                                 |
| Estadístico F               | 624.287*** (df = 71; 68513)        |                                 |
| AIC                         | 112 114.10                         |                                 |

La Tabla presenta los resultados de la regresión por mco. En la columna a la izquierda se muestran los coeficientes y a la derecha el efecto en porcentaje. Entre paréntesis se muestran las desviaciones estándar. \*p<0.1; \*\*p<0.05; \*\*\* p<0.01.

del factor no explicado (coeficientes y variables no observables). Por su parte, 18 pp son explicados por las características iniciales de los individuos tales como entidad federativa, sexo, escolaridad, etcétera. Finalmente, la interacción de ambos factores explica -4 pp de la brecha salarial; esto es, estas interacciones contribuyen a cerrar la brecha. Estos resultados indican que una parte todavía importante de la brecha salarial (6 pp que equivalen a 30% de la brecha total) es atribuible a factores no explicados por el modelo, tales como la discriminación o variables no observables como competencias suaves de las personas.

Figura 3. Descomposición de Oaxaca-Blinder con el modelo lineal



Fuente: Elaboración propia con datos de ENIGH 2018.

Las Figuras presentan el resultado de las variables de mayor interés de la descomposición Oaxaca-Blinder. El Panel (A) muestra la contribución por las diferencias explicadas por características, E; aquella no explicada que se debe a diferencias en los retornos o coeficientes, C, y aquella que se debe a la interacción entre coeficientes y características, CE. El Panel (B) desglosa la contribución de cada una de las variables a los tres componentes presentados en el Panel (A). Se muestran los intervalos de confianza estimados a 95% para 100 réplicas.



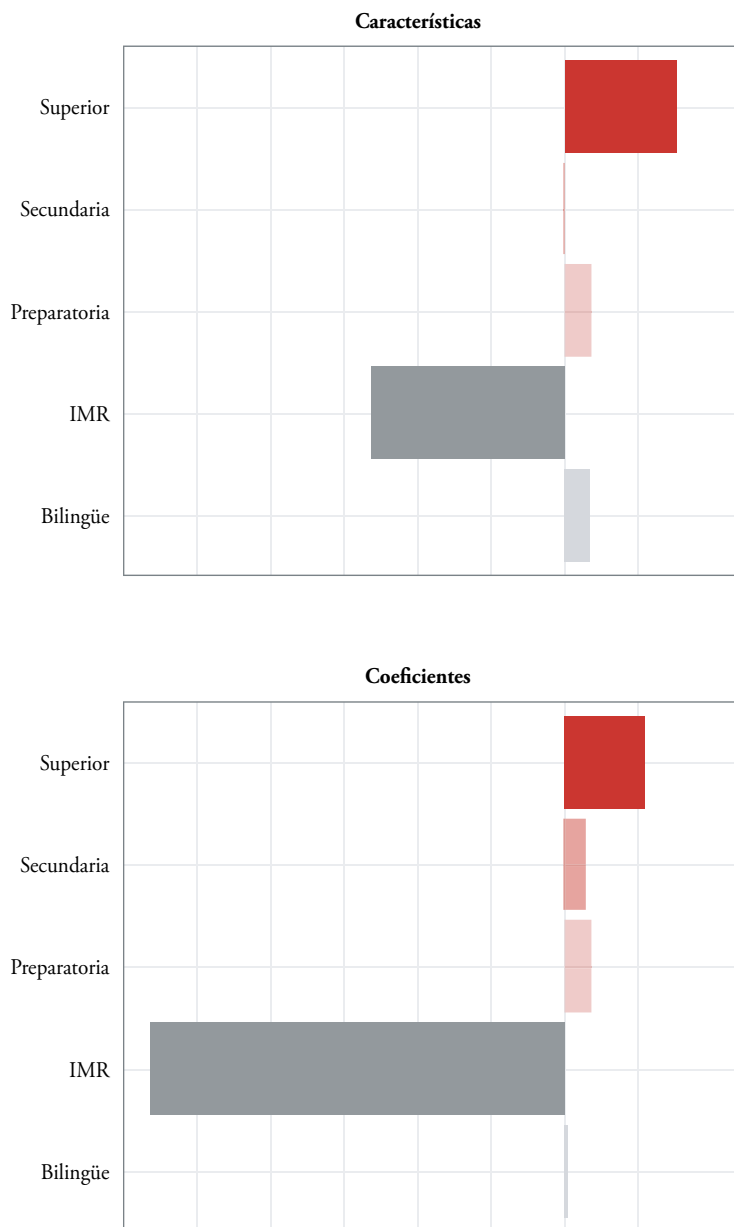
La Tabla 3 presenta los resultados de la estimación del modelo de regresión con cambio de régimen, el cual arroja distintos resultados, respecto de ser bilingüe, que el modelo por MCO, lo cual sugiere la existencia de un sesgo de selección al identificarse como indígena. En particular, el retorno a ser bilingüe no se puede distinguir estadísticamente de cero para ambas subpoblaciones. Los retornos a la educación son crecientes por nivel educativo, tomando como base primaria completa o menos. Igualmente, encontramos que los retornos son mayores para los no indígenas que para los indígenas en todos los niveles educativos. Los signos y la significancia en la covarianza (Rho) muestran que la covarianza del grupo no indígena es estadísticamente significativa, indicando que existe una autoselección en la identificación indígena para este grupo. Esto implica que los pagos en los salarios podrían no ser los mismos si un individuo de este grupo decidiera considerarse indígena. El signo positivo nos indica un sesgo negativo, lo que sugiere que individuos con ingreso mayor a la media tienen una menor probabilidad de identificarse como indígenas. Esto es consistente con los resultados encontrados en el primer modelo y con los hallazgos de Cano-Urbina y Mason (2016) y Canedo (2019).

**Tabla 3: Resultados con el modelo con cambio de régimen**

|                             | Variable dependiente          |                             |                                  |
|-----------------------------|-------------------------------|-----------------------------|----------------------------------|
|                             | <i>Indígena<br/>Selección</i> | <i>No Indígena</i>          | <i>Log(Salario)<br/>Indígena</i> |
| Constante                   | -1.291*(0.094)                | 2.898*** (0.050)            | 2.771*** (0.098)                 |
| Lengua                      | 2.523*** (0.456)              |                             |                                  |
| Bilingüe                    | -0.622 (0.457)                | 0.135 (0.162)               | -0.041 (0.030)                   |
| Secundaria                  |                               | 0.099*** (0.007)            | 0.082*** (0.010)                 |
| Preparatoria                |                               | 0.242*** (0.008)            | 0.209*** (0.013)                 |
| Superior                    |                               | 0.568*** (0.010)            | 0.456*** (0.018)                 |
| MIR                         |                               | -0.144 (0.114)              | 0.070** (0.035)                  |
| <i>Controles:</i>           |                               |                             |                                  |
| <i>Edad</i>                 | Sí                            | Sí                          | Sí                               |
| <i>Mujer</i>                | Sí                            | Sí                          | Sí                               |
| <i>Entidad federativa</i>   | Sí                            | Sí                          | Sí                               |
| <i>Tamaño de localidad</i>  | Sí                            | Sí                          | Sí                               |
| <i>Nivel de escolaridad</i> | Sí                            | Sí                          | Sí                               |
| <i>Sector</i>               | No                            | Sí                          | Sí                               |
| <i>Ocupación</i>            | No                            | Sí                          | Sí                               |
| Sigma                       |                               | 0.542677***                 | 0.567072***                      |
| Rho                         |                               | 0.017176                    | 0.265368***                      |
| R <sup>2</sup>              |                               | 0.977                       | 0.972                            |
| R <sup>2</sup> Ajustada     |                               | 0.977                       | 0.972                            |
| Error. Est.                 |                               | 0.543 (df=48562)            | 0.556 (df=19883)                 |
| Estadístico F               |                               | 28 889.590*** (df=70;48562) | 9,887.991*** (df=70;19883)       |
| AIC                         |                               | 78 694.41                   | 33 265.80                        |
| Log Verosimilitud           |                               |                             | -89 003.18                       |

La Tabla presenta los resultados del modelo con cambio de régimen. La primera columna muestra la regresión de selección; la segunda y tercera columnas muestran los resultados de las variables de interés para los individuos que no se identifican y los que se identifican como indígenas, respectivamente. \*p<0.1; \*\*p<0.05; \*\*\* p<0.01.

Figura 4. Descomposición de Oaxaca en el modelo con cambio de régimen



Fuente: Elaboración propia con datos de ENIGH 2018.

La Figura muestra la contribución a la brecha salarial a cada uno de los componentes de la descomposición de Oaxaca-Blinder de las variables de mayor interés (ecuación 3). IMR es la razón inversa de Mills.

Finalmente, la Figura 4 presenta la contribución de cada una de las variables de interés y sus retornos a la brecha salarial por autoidentificación indígena. Como se puede apreciar, la educación superior continúa siendo una característica que produce brechas salariales, tanto por sus diferencias en la proporción que cuenta con este nivel educativo como por el retorno salarial. Sin embargo, la descomposición elaborada con el modelo con cambio de régimen revela que la brecha se reduce enormemente por la selección de las personas al considerarse indígenas. En la ausencia de autoselección identificación indígena, las brechas serían menores.

Ambos modelos encuentran una brecha por identificación indígena. Una vez que se controla por la identificación indígena de los individuos, los retornos a los distintos niveles educativos presentan pagos diferentes para ambos grupos. Esto es, aquellos individuos que no se consideran indígenas reciben mayores retornos en cada uno de los niveles educativos, tomando como base haber completado solo la primaria o menos. Mientras que el modelo por MCO estimaba un retorno grande y positivo a ser bilingüe, el modelo con cambio de régimen no asigna ningún retorno a esta habilidad, ni para indígenas ni para no indígenas.

## 5. Pruebas de robustez

La evidencia empírica resultante del análisis de los datos en la ENIGH 2018 sugiere la existencia de una brecha salarial debido a la identificación indígena en México. Sin embargo, hay ciertos factores que no podemos analizar a partir de esta encuesta. Por ello, no es posible descartar que exista algún componente inobservable, tales como habilidades, que provoquen estas diferencias en el ingreso. En esta sección analizaremos el efecto de variables observables en nuestras estimaciones y discutiremos sobre el papel potencial de variables no observables en nuestras estimaciones.

### 5.1 Sesgo por selección al empleo remunerado

Nuestra muestra de estimación solo incluye a personas con alguna remuneración laboral. La exclusión de aquellos que no obtienen remuneraciones por su empleo o que no participan en la fuerza laboral puede inducir a sesgos de selección muestral (Heckman, 1974, 1979). Canedo (2019) encuentra que, al corregir por este sesgo de selección muestral, la brecha salarial por origen étnico se reduce de 82.8%

a 77.9%, lo cual es evidencia de que hay selección negativa para entrar al mercado laboral. Aguilar-Rodriguez et al. (2018) hacen la corrección Heckman solo para las mujeres, ya que entre ellas la participación laboral es menor y, por tanto, hay más problemas de sesgo de selección muestral. Sin embargo, no reportan las diferencias en los resultados con y sin corrección.

Como argüimos a lo largo de este trabajo, la autoidentificación indígena también puede inducir a sesgos de selección, y nuestros resultados apuntan a que este es, de hecho, el caso. Es probable que en nuestras estimaciones siga existiendo un sesgo atribuible a la decisión de entrar al mercado laboral remunerado. El problema sería mayor si la decisión de entrar al mercado está correlacionada con la autoidentificación indígena. Existe una amplia literatura en econometría que corrige por dos variables de selección. Sin embargo, la estructura de las decisiones que conllevan a esta selección no se asemeja al problema que tenemos nosotros. Por ejemplo, Fische et al. (1981) modelan la decisión conjunta de las mujeres de asistir a la universidad y trabajar para corregir estimaciones de la ecuación minceriana. Krishnan (1990) utiliza esta misma aproximación para corregir las estimaciones de los salarios y horas trabajadas en un segundo empleo, corrigiendo por la decisión conjunta del esposo de tener ese segundo empleo y de la esposa de tener un trabajo remunerado. Por su parte, Hotchkiss (1991) modela la selección como un problema secuencial, donde primero se determina si se trabaja de tiempo parcial y después se estima el modelo con cambio de régimen de los salarios de trabajadores de tiempo completo y trabajadores de tiempo parcial. Hotchkiss (1991) no corrige por la selección al mercado laboral, como tampoco lo hacen van der Gaag y Vijverberg (1988) en su estimación de los salarios en el sector privado *vs.* público. Lo más cercano a nuestro problema en la literatura es Song (2015), quien modela la decisión de migrar y la decisión del sector de ocupación para después corregir las funciones de salarios por selección. Sin embargo, su muestra solo incluye a quienes ya se encuentran en el mercado laboral. Hacer una doble corrección controlando por doble selección (al mercado laboral e identificación indígena) en un modelo con cambio de régimen está fuera del alcance de este estudio y queda desarrollar este tema como trabajo futuro.<sup>9</sup>

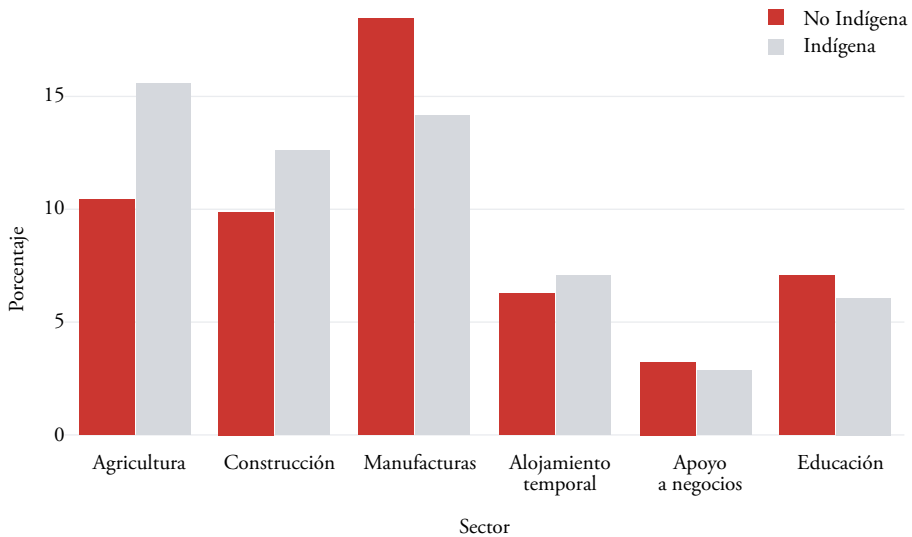
---

<sup>9</sup> Miranday Rabe-Hesketh (2006) estiman modelos con sesgo de selección muestral y cambio de régimen con variables binarias, ordinales y de conteo, los cuales tampoco se adaptan a nuestro problema.

## 5.2 Actividad económica a la que se dedica la empresa y ocupación

Es posible que gran parte de la brecha se deba al tipo de ocupación y al sector de actividad económica a la que se dedica la empresa en la que están empleadas las personas que se identifican como indígenas y las que no.<sup>10</sup> Como muestra la Figura 5 (y el apéndice A.1), el porcentaje de individuos que se consideran indígenas en empresas cuya actividad económica es agrícola, es considerablemente mayor que el de quienes no se identifican como indígenas. En empresas cuya actividad económica es la manufactura, por ejemplo, sucede lo contrario.

Figura 5. Porcentaje de población por sector de actividad económica



Fuente: Elaboración propia con datos de ENIGH 2018 y tomando en cuenta las claves en el Sistema de Clasificación Industrial de América del Norte (SCIAN) 2018.

Se muestra el porcentaje de individuos por identificación indígena en cada grupo de actividad económica. En la Tabla A.1 del apéndice se presenta una lista completa.

Al controlar por la actividad económica a la que se dedica la empresa y la ocupación se puede observar que la brecha salarial disminuye de manera significativa; sin embargo, no desaparece y sigue siendo relativamente alta (-10%). Controlar por la actividad económica reduce el retorno por haber terminado la secundaria

<sup>10</sup> Actividad económica a la que se dedica la empresa de acuerdo con el SCIAN 2018.

en comparación con la primaria en 4 pp, la preparatoria en 10 pp y los niveles superiores en 22 pp. La ocupación los reduce en 5 pp, 15 pp y 37 pp, respectivamente. Finalmente, controlando por ambos se reducen en 6 pp, 18 pp y 47 pp, respectivamente (Tabla 4).<sup>11</sup>

**Tabla 4: Resultados control actividad económica a la que se dedica la empresa y ocupación**

|                         | Variable dependiente: Log(Salario) |            |            |            |
|-------------------------|------------------------------------|------------|------------|------------|
|                         | (1)                                | (2)        | (3)        | (4)        |
| Constante               | 2.314***                           | 2.725***   | 2.371***   | 2.759***   |
| Indígena                | -0.120***                          | -0.118***  | -0.103***  | -0.104***  |
| Lengua                  | -0.572***                          | -0.501***  | 0.447***   | -0.484***  |
| Bilingüe                | 0.460***                           | 0.398***   | -0.544***  | 0.388***   |
| Secundaria              | 0.165***                           | 0.128***   | 0.119***   | 0.104***   |
| Preparatoria            | 0.424***                           | 0.329***   | 0.277***   | 0.240***   |
| Superior                | 1.005***                           | 0.787***   | 0.633***   | 0.549***   |
| Controles:              |                                    |            |            |            |
| <i>Edad</i>             | Sí                                 | Sí         | Sí         | Sí         |
| <i>Mujer</i>            | Sí                                 | Sí         | Sí         | Sí         |
| <i>Sector</i>           | No                                 | Sí         | No         | Sí         |
| <i>Ocupación</i>        | No                                 | No         | Sí         | Sí         |
| Observaciones           | 68 585                             |            |            |            |
| R <sup>2</sup>          | 0.264                              | 0.305      | 0.323      | 0.345      |
| R <sup>2</sup> Ajustada | 0.264                              | 0.305      | 0.323      | 0.344      |
| AIC                     | 125 195.30                         | 121 301.70 | 119 438.90 | 117 282.20 |

La Tabla presenta los resultados para el modelo por mínimos cuadrados ordinarios variando controles de ocupación y sector de ocupación. \*p<0.1; \*\*p<0.05; \*\*\* p<0.01.

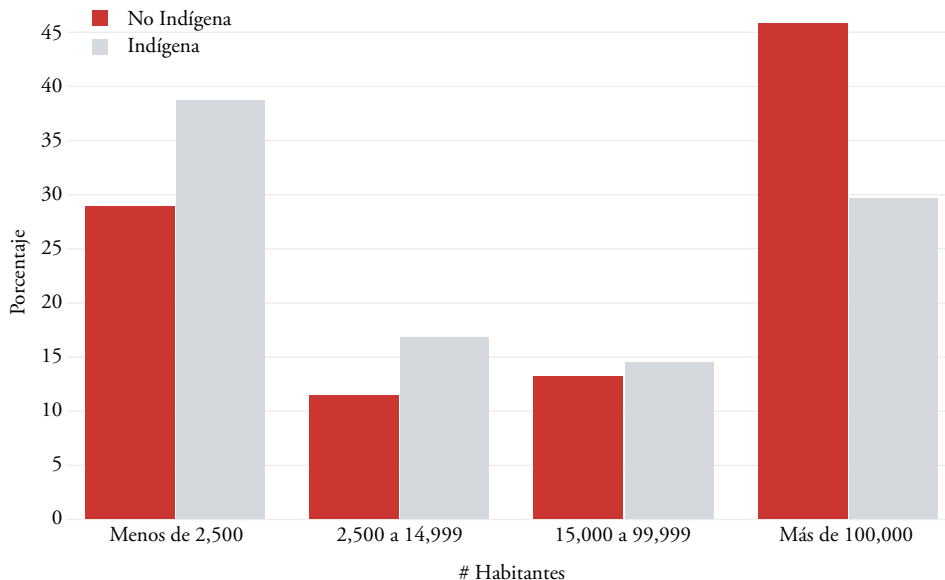
### 5.3 Estado y tamaño de localidad

Como se mencionó en la primera parte de este trabajo, la mayor parte de la población indígena se encuentra concentrada en ciertos estados del país (Figura 1). Asimismo, una gran proporción de trabajadores indígenas se ubica en localidades con menos de 15 mil habitantes (Figura 6). Las columnas (2), (3) y (4)

<sup>11</sup> La Tabla 6 presenta esta y el resto de las pruebas de robustez de nuestras estimaciones con el modelo con cambio de régimen.

de la Tabla 5 introducen controles por tamaño de localidad y entidad federativa. Controlar por entidad federativa reduce la brecha salarial por identificación indígena en casi 6 pp (columna 3 *vs.* columna 1), mientras que controlar solo por el tamaño de la localidad la reduce en 1 pp. Ambos controles reducen la brecha en 6.2 pp. Por su parte, la brecha salarial por hablar lengua indígena también se reduce en mayor medida cuando se controla por estado (-11 pp, columna 3 *vs.* 1) que cuando se controla por tamaño de la localidad (-0.8 pp columna 2 *vs.* 1). Finalmente, los retornos a la educación son relativamente robustos a la inclusión de estos dos controles adicionales.

**Figura 6. Porcentaje de población por tamaño de localidad**



Fuente: Elaboración propia con datos de ENIGH 2018.  
En la Tabla A.2 del apéndice se presenta una lista completa.

Estos resultados pueden ser explicados puesto que más de la mitad de los individuos que se identifican como indígenas habitan en comunidades de menos de 15 mil habitantes, mientras que más de la mitad de quienes no se identifican viven en comunidades de más de 15 mil habitantes (Figura 6 y Tabla A.2 del apéndice).



Esto puede implicar problemas de oportunidades laborales en todo el territorio. Asimismo, existe una relación fuerte entre el sector en el que se desempeñan los individuos y el estado donde habitan.

#### **5.4 Sesgos de selección por migración**

Otra posible explicación es que la emigración induzca a sesgos de selección en nuestras estimaciones. Esto se soporta ya que, anteriormente, encontramos que gran parte de la brecha salarial que existe para quienes se identifican como indígenas es explicada por el tamaño de la localidad y el estado en el que habitan. Desafortunadamente, la ENIGH no pregunta el lugar de nacimiento de las personas ni tampoco identifica cambios en la autoadscripción indígena. Sin embargo, aquí discutimos los mecanismos a través de los cuales la migración induce sesgo.

Primero, es necesario hacer notar la posibilidad de que al migrar las personas tiendan a subreportar la identificación indígena. Si migran a una comunidad donde una menor cantidad de personas se consideran indígenas y existe una mayor tendencia a la discriminación (ya sea en el ámbito laboral o social), es probable que algunos individuos no se autoidentifiquen como indígenas, aunque en un pasado sí lo hayan hecho. Esto es parte del proceso de aculturación y asimilación a la cultura dominante. Teniendo esto claro, hay dos posibles mecanismos en acción que explicamos a continuación.

El primer mecanismo es el educativo. Las regiones con población mayoritariamente indígena tienen menos acceso a oferta educativa de niveles superiores. Por ello, aquellos que desean estudiar más allá del nivel básico se pueden ver obligados a migrar a zonas donde sí encuentran esta oferta educativa. Estos individuos tenderán a tener un mayor potencial académico, mayores aspiraciones o, simplemente, una familia de origen más aventajada. Si, además, tienden a no regresar a sus lugares de origen, entonces esta selección positiva a la migración induciría un sesgo en la brecha salarial por identificación indígena.

El segundo mecanismo surge por migración económica. Las regiones donde hay mayor concentración de población indígena tienden a ofrecer menos oportunidades laborales, independientemente del nivel educativo. Dado esto, algunas personas se pueden ver obligadas a migrar en busca de mejores oportunidades de trabajo. Aquellos que se quedan en sus lugares de origen podrían estar seleccionados negativamente por la vía de una menor educación o menores recursos para migrar. De ser este el caso, la migración también tendería a acrecentar la brecha salarial por identificación indígena.

**Tabla 5: Resultados estado y tamaño de localidad**

|                            | Variable dependiente: Log(Salario) |            |            |            |
|----------------------------|------------------------------------|------------|------------|------------|
|                            | (1)                                | (2)        | (3)        | (4)        |
| Constante                  | 2.759***                           | 2.844***   | 2.824***   | 2.887***   |
| Indígena                   | -0.104***                          | -0.093***  | -0.048***  | -0.042***  |
| Lengua                     | -0.484***                          | -0.478***  | -0.372***  | -0.358***  |
| Bilingüe                   | 0.388***                           | 0.380***   | 0.313***   | 0.299***   |
| Secundaria                 | 0.104***                           | 0.096***   | 0.101***   | 0.094***   |
| Preparatoria               | 0.240***                           | 0.223***   | 0.248***   | 0.232***   |
| Superior                   | 0.549***                           | 0.522***   | 0.566***   | 0.542***   |
| 15 000 a 99 999 habitantes |                                    | -0.099***  |            | -0.057***  |
| 2500 a 14 999 habitantes   |                                    | -0.134***  |            | -0.064***  |
| Menos de 2500 habitantes   |                                    | -0.121***  |            | -0.114***  |
| <i>Controles:</i>          |                                    |            |            |            |
| <i>Edad</i>                | Sí                                 | Sí         | Sí         | Sí         |
| <i>Mujer</i>               | Sí                                 | Sí         | Sí         | Sí         |
| <i>Sector</i>              | Sí                                 | Sí         | Sí         | Sí         |
| <i>Ocupación</i>           | Sí                                 | Sí         | Sí         | Sí         |
| <i>Estado</i>              | No                                 | No         | Sí         | Sí         |
| Observaciones              | 68,585                             |            |            |            |
| R <sup>2</sup>             | 0.345                              | 0.351      | 0.389      | 0.393      |
| R <sup>2</sup> Ajustada    | 0.344                              | 0.350      | 0.389      | 0.392      |
| AIC                        | 117,282.20                         | 116,671.70 | 112,496.80 | 112,114.10 |

La Tabla presenta los resultados para el modelo por mínimos cuadrados ordinarios variando controles y desagregando por tamaño de localidad. \*p<0.1; \*\*p<0.05; \*\*\* p<0.01.

Aunque no podemos dar evidencia directa de ninguno de estos mecanismos, las estimaciones por MCO presentadas en el apartado 5.3 muestran que la introducción de controles del tamaño de localidad y entidad federativa reduce la brecha. Las estimaciones del modelo con cambio de régimen muestran que la brecha se reduce solo hasta que se introduce el tamaño de localidad como control (Tabla 6), y hay ciertas reducciones en las brechas por retornos cuando se controla por entidad federativa o tamaño de localidad. Los controles por ocupación y sector reducen las brechas de los retornos en el nivel preparatoria y superior, lo cual sugiere que también hay selección a las ocupaciones y el sector donde se labora.

Sin embargo, añadir variables de migración, ocupación o sector podrían generar problemas de endogeneidad adicionales en el modelo que terminen por invalidar las estimaciones.

### **5.5 Otros factores**

Existen otros factores no observables que pueden estar relacionados con las variables de interés, tales como la fluidez en el español, el tono de piel, o el nombre y apellido del individuo. Otros trabajos tales como Cano-Urbina y Mason (2016), Rojas Lomelín et al. (2019), Arceo-Gómez y Campos-Vázquez (2014) y Campos Vázquez et al. (2017) encuentran que estos factores también contribuyen a las brechas en resultados laborales.

En la base de datos no podemos observar si un individuo habla el español con fluidez o no. Dado que es el idioma dominante en México, hablar esta lengua con menor fluidez o acentos indígenas podría disparar estereotipos negativos o discriminación. Asimismo, el manejo de la lengua puede estar relacionado con ciertas habilidades sociales requeridas en el mercado laboral. Igualmente, no es posible controlar por el tono de piel con los datos de ENIGH 2018, pero sabemos que quienes se identifican como indígenas tienden a tener tonalidades más oscuras. Del Módulo de Movilidad Social Intergeneracional 2016 del INEGI se reveló que el tono de piel se encuentra fuertemente correlacionado con el nivel de riqueza.

**Tabla 6. Pruebas de robustez con el modelo de regresión con cambio de régimen**

|                            | Variable dependiente: Log(Salario) |                     |                      |                     |                     |                     |                     |                     |                     |                     |
|----------------------------|------------------------------------|---------------------|----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                            | (1)                                |                     | (2)                  |                     | (3)                 |                     | (4)                 |                     | (5)                 |                     |
|                            | <i>N.I.</i>                        | <i>I.</i>           | <i>N.I.</i>          | <i>I.</i>           | <i>N.I.</i>         | <i>I.</i>           | <i>N.I.</i>         | <i>I.</i>           | <i>N.I.</i>         | <i>I.</i>           |
| Constante                  | 2.452***<br>(0.043)                | 1.981***<br>(0.07)  | 2.579***<br>(0.045)  | 1.811***<br>(0.089) | 2.460***<br>(0.049) | 2.232***<br>(0.099) | 2.472***<br>(0.048) | 2.245***<br>(0.096) | 2.898***<br>(0.050) | 2.771***<br>(0.098) |
| Bilingüe                   | 0.483***<br>(0.048)                | 0.110***<br>(0.017) | 1.111***<br>(0.074)  | 0.152***<br>(0.028) | 0.176<br>(0.172)    | -0.001<br>(0.032)   | 0.178<br>(0.165)    | -0.009<br>(0.031)   | 0.135<br>(0.162)    | -0.041<br>(0.030)   |
| Secundaria                 | 0.162***<br>(0.007)                | 0.150***<br>(0.011) | 0.143***<br>(0.007)  | 0.132***<br>(0.011) | 0.141***<br>(0.007) | 0.125***<br>(0.011) | 0.108***<br>(0.007) | 0.096***<br>(0.010) | 0.099***<br>(0.007) | 0.082***<br>(0.010) |
| Preparatoria               | 0.432***<br>(0.008)                | 0.394***<br>(0.013) | 0.401***<br>(0.008)  | 0.378***<br>(0.013) | 0.396***<br>(0.008) | 0.360***<br>(0.013) | 0.275***<br>(0.008) | 0.246***<br>(0.013) | 0.242***<br>(0.008) | 0.209***<br>(0.013) |
| Superior                   | 1.009***<br>(0.008)                | 0.964***<br>(0.015) | 0.980***<br>(0.008)  | 0.962***<br>(0.015) | 0.975***<br>(0.009) | 0.936***<br>(0.015) | 0.652***<br>(0.010) | 0.562***<br>(0.018) | 0.568***<br>(0.010) | 0.456***<br>(0.018) |
| IMR                        | -0.353***<br>(0.013)               | 0.270***<br>(0.013) | -0.857***<br>(0.043) | 0.350***<br>(0.30)  | -0.182<br>(0.12)    | 0.138***<br>(0.037) | -0.176<br>(0.120)   | 0.110***<br>(0.036) | -0.144<br>(0.114)   | 0.070***<br>(0.035) |
| <i>Controles:</i>          |                                    |                     |                      |                     |                     |                     |                     |                     |                     |                     |
| <i>Edad</i>                | Sí                                 |                     | Sí                   |                     | Sí                  |                     | Sí                  |                     | Sí                  |                     |
| <i>Mujer</i>               | Sí                                 |                     | Sí                   |                     | Sí                  |                     | Sí                  |                     | Sí                  |                     |
| <i>Entidad federativa</i>  | No                                 |                     | Sí                   |                     | Sí                  |                     | Sí                  |                     | Sí                  |                     |
| <i>Tamaño de localidad</i> | No                                 |                     | No                   |                     | Sí                  |                     | Sí                  |                     | Sí                  |                     |
| <i>Sector</i>              | No                                 |                     | No                   |                     | No                  |                     | Sí                  |                     | Sí                  |                     |
| <i>Ocupación</i>           | No                                 |                     | No                   |                     | No                  |                     | No                  |                     | Sí                  |                     |
| Sigma                      | 0.63 ***                           | 0.67***             | 0.59***              | 0.64***             | 0.57***             | 0.61***             | 0.55***             | 0.59***             | 0.54***             | 0.56***             |
| Rho                        | 0.53 ***                           | 0.52 ***            | 0.38***              | 0.48***             | 0.02                | 0.35***             | 0.01                | 0.33***             | 0.01                | 0.26***             |

Fuente: Elaboración propia con datos de ENIGH 2018.

La Tabla presenta los resultados para el modelo con cambio de régimen variando controles. Los errores estándar robustos se muestran entre paréntesis. \*p&lt;0.1; \*\*p&lt;0.05; \*\*\* p&lt;0.01.

## 6. Conclusiones

México es un país altamente desigual y el origen étnico-racial es una de las dimensiones que contribuyen a esta desigualdad. Existen algunos estudios que han analizado las diferencias salariales por origen-étnico racial y rasgos lingüísticos, pero todavía hace falta mucha literatura al respecto en las ciencias económicas. De nuestra revisión encontramos que hasta ahora ningún estudio ha reconocido que la autoidentificación indígena puede llevar a sesgos de selección en el análisis. Nuestra premisa es que la identidad indígena no se puede tomar como algo estático o heredado. Las personas modifican su identidad a lo largo de la vida de acuerdo con distintos propósitos, tales como conformación a la identidad dominante en la cultura, elusión de la discriminación, entre otros procesos intencionales o no intencionales de aculturación. En este artículo resolvemos este problema mediante la estimación de un modelo con cambio de régimen, el cual estima, en una primera etapa, la selección en la autoidentificación y, en una segunda, dos ecuaciones de los salarios según cómo se autoidentificó el trabajador.

Usando los datos de ENIGH 2018, encontramos que en México existe una brecha en salarios asociada a la identificación indígena. En específico, podemos observar que existe una diferencia no ajustada de 23% en los salarios: la población que se considera indígena gana 23% menos que la población que no se considera así. De acuerdo con la descomposición de Oaxaca en el modelo por MCO, encontramos que 70% de esta brecha se debe a diferencias en las características de los indígenas en comparación con los no indígenas. En particular, los indígenas todavía tienen un rezago en la escolaridad, habitan en estados con peores salarios, y trabajan en ocupaciones menos remuneradas. Así, la brecha que se debe a características se puede cerrar con el acceso a mayores oportunidades educativas y laborales para la población indígena.

El modelo estimado por MCO encuentra una fuerte asociación negativa entre hablar una lengua nativa en comparación con quien no la habla, y un pago adicional a quienes son bilingües en comparación con quienes solo hablan español. Sin embargo, en el modelo con cambio de régimen, los pagos por ser bilingüe son nulos, tanto para personas que se identifican como indígenas como las que no lo hacen: hablar una lengua indígena además de español no supone un retorno en el mercado laboral. Encontramos también una selección negativa en la autoidentificación indígena, la cual indica que quienes se declaran indígenas tienen salarios más bajos que un individuo aleatorio de la muestra. La corrección

por selección desempeña un papel importante en descomposición de Oaxaca, pero aun después de corregirla las características continúan explicando parte de la brecha salarial por identificación étnica. Una gran mayoría de la diferencia que encontramos en los salarios también es explicada por el estado y el tamaño de localidad donde viven. Los datos sugieren que los individuos que se identifican como indígena están predominantemente en sectores agrícolas y de construcción, que se desarrollan sobre todo en el sur del país, la región con mayores rezagos.

Una limitante de nuestro análisis es que no hemos corregido por selección de participación en el mercado laboral, pues esto complicaría seriamente la estimación de nuestro modelo. Esto puede ser un problema serio, ya que un alto porcentaje de quienes se declaran indígenas trabajan sin remuneraciones. Dado que, en general, la literatura encuentra una selección positiva al empleo entre grupos minoritarios (como las mujeres), es de esperarse que la brecha salarial sería todavía mayor si todos tuviesen un empleo remunerado. Dejamos este asunto pendiente para futuras investigaciones.

## Referencias

- Adamchik, V. A. y Bedi, A. S. (2000). Wage differentials between the public and the private sectors: evidence from an economy in transition. *Labour Economics*, 7(2), 203-224.
- Advani, A., Ash, E., Cai, D. y Rasul, I. (2021). Race-related research in economics and other social sciences. CEPR Discussion Paper No. DP16115.
- Aguilar-Rodriguez, A., Miranda, A. y Zhu, Y. (2018). Decomposing the language pay gap among the indigenous ethnic minorities of Mexico: Is it all down to observables? *Economics Bulletin*, 38(2), 689-695.
- Arceo-Gómez, E. O. y Campos-Vázquez, R. M. (2014). Race and marriage in the labor market: A discrimination correspondence study in a developing country. *American Economic Review*, 104(5), 376-380.
- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4), 436-455.
- Campos-Vázquez, R. M. (2021). The higher price of whiter skin: an analysis of escort services. *Applied Economics Letters*, 28(1), 1-4.
- Campos Vázquez, R. M. et al. (2017). Identidad social y estereotipos por color de piel. Aspiraciones y desempeño en jóvenes mexicanos. *El Trimestre Económico*, 85(337), 53-79.

- 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), 12-27.
- Cano-Urbina, J. y Mason, P. L. (2016). Acculturation and the labor market in Mexico. *IZA Journal of Labor Policy*, 5(1), 1-29.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. (29 de enero de 2021). *Pobreza en la población indígena*. [https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza\\_Indigena.aspx](https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza_Indigena.aspx)
- El Colegio de México (2018). *Desigualdades en México 2018*. Ciudad de México: El Colegio de México, Red de Estudios sobre Desigualdades 2018.
- Fishe, R. P. H., Trost, R. P. y Lurie, P. M. (1981). Labor force earnings and college choice of young women: An examination of selectivity bias and comparative advantage. *Economics of Education Review*, 1(2), 169-191.
- Gutiérrez Chong, N. (2012). *Mitos nacionalistas e identidades étnicas: los intelectuales indígenas y el Estado mexicano*. Ciudad de México: Instituto de Investigaciones Sociales, UNAM.
- Heckman, J. (1974). Shadow primes, market wages, and labor supply. *Econometrica*, 42(4), 679-694.
- Heckman, J. (1979). Sample selection bias a specification error. *Econometrica*, 47(1), 153-162.
- Hisamatsu, Y. y Ukeda, H. (2002). Economic disadvantage of being indigenous: An analysis of Mexico's first *National Survey of Employment in Indigenous Areas*. Prepared for delivery at the 2003 meeting of the Latin American Studies Association, Dallas, Texas, Marzo 27-29, 2003.
- Horbath, J. E. (2008). La discriminación laboral de los indígenas en los mercados urbanos de trabajo en México: revisión y balance de un fenómeno persistente. *Pobreza, exclusión social y discriminación étnico-racial en América Latina y el Caribe*, 25, 25-52.
- Hotchkiss, J. L. (1991). The definition of part-time employment: A switching regression model with unknown sample selection. *International Economic Review*, 32(4), 899-917.
- Krishnan, P. (1990). The economics of moonlighting: A double self-selection model. *Review of Economics and Statistics*, 72(2), 361-367.
- Lee, L. F. (1978). Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review*, 19(2), 415-433.

- Miranda, A. y Rabe-Hesketh, S. (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *The Stata Journal*, 6(3), 285-308.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3), 693-709.
- Robinson, C. y Tomes, N. (1982). Union wage differentials in the public and private sectors: A simultaneous equations specification. *Journal of Labor Economics*, 2(1), 106-127.
- Rojas Lomelín, M. A., Aguad Revilla, J. y Morrison, J. A. (2019). *Diversidad étnico-racial en México y su influencia en la movilidad social*. Inter-American Development Bank.
- Solís, P., Krozer, A., Arroyo Batista, C. y Güémez Graniel, B. (2019a). Discriminación étnico-racial en México: una taxonomía de las prácticas. En J. Rodríguez Zepeda y T. González Luna Corvera (coords.), *La métrica de lo intangible: del concepto a la medición de la discriminación* (pp. 55-94). Ciudad de México: Consejo Nacional para Prevenir la Discriminación.
- Solís, P., Güémez Graniel, B. y Lorenzo Holm, V. (2019b). *Por mi raza hablará la desigualdad: Efectos de las características étnico-raciales en la desigualdad de oportunidades en México*. Oxfam México, Informe Técnico.
- Song, Y. (2015). Hukou-based labour market discrimination and ownership structure in urban China. *Urban Studies*, 53(8), 1657-1673.
- Suárez Blanch, C. (2006). Situación sociodemográfica de los pueblos indígenas de México. En F. Del Popolo, y M. Ávila (eds.). *Pueblos indígenas y afrodescendientes de América Latina y el Caribe: información sociodemográfica para políticas y programas* (pp. 85-95) Santiago de Chile: CEPAL.
- Van der Gaag, J. y Vijverberg, W. (1988). A switching regression model for wage determinants in the public and private sectors of a developing country. *Review of Economics and Statistics*, 70(2), 244-252.
- Wooldridge, J. (2010). *Introducción a la Econometría*, 4a. ed. Cengage Learning.



## Apéndice

**Tabla A.1: Porcentaje por actividad económica**

|  | Identificación     |                 |
|--|--------------------|-----------------|
|  | <i>No indígena</i> | <i>Indígena</i> |
| Actividades legislativas               | 0.058              | 0.053           |
| Agricultura                            | 0.105              | 0.156           |
| Comercio al por mayor                  | 0.033              | 0.029           |
| Comercio al por menor                  | 0.099              | 0.092           |
| Construcción                           | 0.099              | 0.126           |
| Corporativos                           | 0                  | 0               |
| Energía, agua y gas                    | 0.006              | 0.005           |
| Información en medios masivos          | 0.008              | 0.005           |
| Manufacturas                           | 0.185              | 0.142           |
| Minería                                | 0.009              | 0.006           |
| Otros servicios                        | 0.099              | 0.122           |
| Alojamiento temporal                   | 0.063              | 0.071           |
| Apoyo a negocios                       | 0.033              | 0.029           |
| Esparcimiento cultural                 | 0.007              | 0.008           |
| Servicios de salud y asistencia social | 0.039              | 0.028           |
| Educación                              | 0.071              | 0.061           |
| Servicios financieros                  | 0.012              | 0.006           |
| Servicios inmobiliarios                | 0.006              | 0.007           |
| Profesionales y técnicos               | 0.019              | 0.009           |
| Transportes y almacenamiento           | 0.049              | 0.045           |

Fuente: Elaboración propia con datos de ENIGH 2018 y tomando en cuenta las claves en el SCIAN 2018. Se muestra el porcentaje de individuos por identificación indígena en cada grupo de actividad económica.

**Tabla A.2. Porcentaje por tamaño de localidad**

|                            | No indígena       |                  | Indígena          |                  |
|----------------------------|-------------------|------------------|-------------------|------------------|
|                            | <i>Frecuencia</i> | <i>Acumulado</i> | <i>Frecuencia</i> | <i>Acumulado</i> |
| Más de 100 000 habitantes  | 45.97             | 45.97            | 29.84             | 29.84            |
| 15 000 a 99 999 habitantes | 13.34             | 59.31            | 14.57             | 44.41            |
| 2500 a 14 999 habitantes   | 11.62             | 70.93            | 16.84             | 61.25            |
| Menos de 2500 habitantes   | 29.07             | 100              | 38.75             | 100              |

La Tabla presenta el porcentaje de individuos que habitan en cada tamaño de localidad por identificación indígena.