

Universidad Nacional de Colombia

Facultad de Ciencias Económicas - Econometría I

Trabajo Final

Luis Ángel Palacio Bayuelo G01
Augusto David Rico Dautt G01

1. Introducción

Una parte importante para verificar el estado de una economía se encuentra en la evolución de su mercado de trabajo, específicamente el nivel salarial, que está íntimamente ligado al nivel educativo. Gran parte de las teorías existentes coinciden en considerar la educación como una variable de inversión que aumenta la productividad de las personas y la probabilidad de que un individuo perciba salarios más elevados. Sin embargo la teoría como la práctica, demuestran que a parte de la educación, existen otras características como el género, la edad, la experiencia, la habilidad innata, pueden ser cruciales para que una persona consiga un puesto de trabajo bien remunerado.

Para el trimestre enero - marzo de 2022, la tasa de desempleo nacional fue 13,2 %, la tasa global de participación se ubicó en 63,4 %, y la tasa de ocupación en 55,0 %, como parados con los datos del año 2019, en el que la tasa de desempleo fue 9,5 %, la tasa global de participación se ubicó en 63,6 %, y la tasa de ocupación se ubicó en un 58,6 %. Es evidente que la tasa de desempleo subió, a causa de la crisis sanitaria covid 19. Desde el último siglo los ingresos salariales en Colombia han sido crecientes, sin embargo persisten problemas estructurales en cuanto al ajuste salarial, tomando en cuenta el contexto económico del país.

El objetivo de este trabajo, es mostrar la relación que existe entre el salario de los colombianos con los años de educación, la experiencia del trabajador, la antigüedad en el puesto de trabajo, las horas remuneradas, las horas no remuneradas, la etnia, el analfabetismo, el género, la sindicalización del trabajador y si tiene un segundo trabajo. Para el análisis de esta relación, se utilizó el modelo salarial de Mincer (1974) y sus variantes: modelo simple y modelo extendido, usando como fuente de información la Gran Encuesta Integrada de Hogares (GEIH) del Departamento Administrativo Nacional de Estadística (DANE) para diciembre de 2021.

2. Objetivos

- Evidenciar la relación que existe entre el salario de los colombianos con los años de educación, la formalidad, la antigüedad en el puesto de trabajo, la experiencia, el género, la sindicalización del trabajador y la edad.

3. Marco Teórico

La idea principal de considerar a la educación como una forma de inversión, ligada a la teoría del capital humano, viene de uno de los pensadores más grande en economía, Adam Smith. Smith resalta que Un hombre educado a costa de mucho trabajo y tiempo, puede ser comparado con una máquina costosa. El trabajo que aprenda a realizar más allá de los salarios habituales de la mano de obra corriente, repondrá todo el gasto de su educación (Smith,1776). En el párrafo anterior se deja claro la importancia de la educación en la determinación de los salarios corrientes y futuros entre diferentes tipos de trabajadores.

Solo hasta mediados de los años cincuenta y principio de los sesenta se empezaron a desarrollar modelos que tuvieron en cuenta la inversión en educación como una decisión que los individuos toman pensando en incrementar sus ingresos en un

futuro, la utilidad y el bienestar de las personas.

Uno de los modelos más utilizados por los economistas para estimar el impacto de la educación sobre los salarios es el modelo de Jacob Mincer (1974). Entre las características que determinan los salarios, se pueden encontrar las capacidades adquiridas como consecuencia de la inversión en capital humano (educación formal, experiencia laboral), capacidades debidas a habilidades innatas del individuo (“inteligencia”), tipo de empleo (rama de actividad, por ejemplo), género, edad y otras como la pertenencia a un sindicato o la formalidad.(Bour, Sf)

La ecuación tradicional de Mincer (1974), estima por mínimos cuadrados ordinarios (MCO) los coeficientes de un modelo semi-logarítmico, usando como variable dependiente el logaritmo de los ingresos y como variables independientes los años de educación, la experiencia laboral y el cuadrado de ésta.(Friere y Tejeiro, Sf). El modelo general viene dado por:

$$\ln(W) = \beta_0 + \beta_1 EDU + \beta_2 EXP + \beta_3 EXP^2 + u_i$$

Donde, W son los ingresos del individuo; EDU es el número de años de educación formal completada; EXP son los años de experiencia laboral; U_i es un término de perturbación aleatoria que se distribuye según Normal $(0, \sigma_{u_i}^2)$

La función de ingresos de Mincer postula el paralelismo de los perfiles del logaritmo del ingreso con respecto a los distintos niveles de educación (Sapelli, 2003, citado por Friere y Tejeiro, Sf), en este modelo generalizado el valor del coeficiente de los años de educación formal se interpreta como la tasa de rendimiento media de un año adicional de estudio poseído por los trabajadores. Por otro lado, tomando en cuenta la teoría de los perfiles edad-ingresos, se tiene que conforme aumenta la experiencia, los ingresos individuales aumentan, pero cada año de experiencia tiene un efecto sobre los ingresos menor que el anterior, por tanto se espera que al ser la función cóncava con relación a la experiencia, la estimación de β_2 sea positiva y la de β_3 sea negativa. (Friere y Tejeiro, Sf)

Las simplificadoras del modelo de Mincer se resumen en:

- Los costos de la inversión en educación son, únicamente, costos de oportunidad (ingresos que dejan de percibirse.
- el tiempo que un individuo permanece en el mercado laboral es independiente del nivel de estudios alcanzado, y se supone que su permanencia en el mismo es continua
- los individuos comienzan a trabajar al finalizar sus estudios.

La función de ingresos minceriana ha sido sinónimo de críticas por aquellos que han analizado su forma funcional, su crítica esta orientada a la estructura simple que no logra caracterizar completamente la distribución conjunta del ingreso, la edad y la escolaridad. A pesar de esto, muchos economistas siguen usando este modelo debido a que es un buen punto de partida para la realización de modelos con un mayor índice de complejidad.

El primer problema del modelo se encuentra en suponer rendimientos a la escolaridad idénticos para toda la población. Los costos marginales que enfrenta cada persona para asistir a la escuela no son iguales. Cada individuo enfrenta una restricción monetaria diferente, por tanto la decisión de adquirir un año más de educación es diferente para cada individuo, además a la restricción monetaria se le puede sumar la capacidad innata que tiene los individuos que facilita el proceso de aprendizaje.

Un segundo problema del modelo de Mincer es especificar las variables educación y experiencia laboral como de tiempo continuo, ya que en la practica estas variables se usan en tiempo discreto. Estos problemas han llevado al mejoramiento de los modelos de estimación. Se ha sugerido utilizar una serie de variables dummy que ayuden a complementar y entender la estimación salarial.

4. Presentación del modelo económico

Basándose en el modelo de Jacob Mincer de salarios, se considero apropiado utilizar el minceriano simple y adicionalmente considerar aspectos económicos como el origen étnico, género, horas trabajadas remuneradas así como horas trabajadas no remuneradas, si la persona se encuentra sindicalizada , el tiempo que lleva trabajando en el lugar de trabajo y si tiene un segundo trabajo, estas variables se consideraran significantes para determinar el valor de los salarios por los impactos que estas pueden tener en estabilidad laboral, las posibilidades para poder ejercer una jornada completa e incluso la posibilidad de tener acceso a una mejor educación.

Hay una variable adicional que se eligió que es el analfabetismo para poder comprender si realmente ser analfabeta, que en teoría es una persona con cero o excesivamente pocos años de estudio, afecta en la determinación de su salario, si el modelo de Mincer se cumple esta variable debe ser fuertemente negativa, no obstante también puede suceder que personas analfabetas no logren integrarse al tejido laboral y por ende la muestra de personas analfabetas que trabajan no es representativa o se encuentre sesgada.

Se eligió analizar el origen étnico debido a que muchas de estas comunidades tienen múltiples problemas de educación, así como de integración al tejido laboral, debido a su origen y cultura, aunque la razón principal es porque los territorios donde nacieron no son centros de empleo y muchas de estas zonas son marginadas económicamente; como puede ser el pacífico colombiano para los afro-descendientes o la selva amazónica para los indígenas, zonas características por la alta informalidad y pocas industrias.

Se consideraron temas como las horas no remuneradas, ya que estas en su mayoría son tareas que se les encarga a las mujeres y estas labores del hogar impactan directamente en la capacidad de poder ejercer un empleo de tiempo completo, pero aún mas importante en la capacidad de poder utilizar el tiempo fuera del trabajo para poder estudiar y mejorar la propia capacidad productiva, por lo que se espera que esta sea una variable que impacte negativamente el salario.

Se consideró pertinente evaluar si la persona se encuentra sindicalizada, ya que esta variable tiende a ser un sinónimo de estabilidad laboral, mejores condiciones y salarios más altos o si por el contrario la persona puede estar precarizada laboralmente, lo que la obliga a buscar otro trabajo, se espera que la hora laborada de un sindicalizado sea mayor, también se puede esperar que una persona con segundo empleo tenga un impacto negativo en su salario, esta misma razón, implica que la variable de horas remuneradas genere un impacto negativo por la relación que puede tener una persona precarizada, quienes comúnmente tienen menores salarios y mayor cantidad de horas trabajadas a unos costos menores.

Por último, el género se analiza para comprender si realmente existe un impacto en la determinación del salario por el hecho de ser mujer, en un escenario ideal esta variable no debería ser significativa y su impacto debería ser neutro no obstante por la realidad nacional se cree que esta variable puede tener un impacto negativo en la determinación de los salarios.

5. Presentación de los datos

El modelo Minceriano que se utilizó para la estimación de los salarios es un modelo extendido que se define de la siguiente manera

$$\ln(W) = \beta_0 + \beta_1 EDU + \beta_2 EXP + \beta_3 EXP^2 + \beta_4 EXPT + \beta_5 hr + \beta_6 hnr + \delta D_i + \delta D_a + \delta D_{anfb} + \delta D_g + \delta D_s + \delta D_{st} + u_i$$

Donde la variable endógena $\ln W$, representa el logaritmo del salario por hora del individuo, la variables exógenas EDU es número de años que lleva estudiando, EXP es la experiencia, $EXPT$ es tiempo que lleva el individuo trabajando en el mismo, lugar, negocio o empresa, hr representa el número de horas remuneradas, hnr el número de horas no remuneradas por economía del hogar, y la variables dummies D_i si la persona pertenece a una comunidad indígena, D_a si pertenece a la población afrocolombiana, D_{anfb} si la persona es analfabeta, D_g el género, D_s si el individuo esta o no sindicalizado, por último la variable dummy D_{st} establece si el individuo tiene un segundo trabajo.

A manera de resumen, las variables incluidas en los anteriores datos se encuentran en el siguiente cuadro:

| Variables de Interés | Dummy | Grupo de tratamiento D=1 |
|----------------------|------------|---------------------------------------|
| Indígena | D_i | Si pertenece a una comunidad indígena |
| Afro-descendiente | D_a | Si es afro-descendiente |
| Analfabetismo | D_{anfb} | Si es analfabeta |
| género | D_g | Si es mujer |
| Sindicalizado | D_s | Si esta sindicalizado |
| Segundo trabajo | D_{st} | Si tiene un segundo trabajo |

En el siguiente cuadro se presentan las principales medidas de tendencia central: media y cuartiles; y medidas de dispersión: varianza y desviación estándar de las variables exógenas cuantitativas.

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | Var | SD |
|----------------------|-------|---------|--------|-------|---------|--------|--------|-------|
| Educacion | 0.00 | 9.00 | 11.00 | 10.86 | 13.00 | 26.00 | 68.52 | 8.28 |
| Experiencia | -3.00 | 10.00 | 21.00 | 23.20 | 34.00 | 78.00 | 0.01 | 0.11 |
| Tiempo Empresa | 0.00 | 0.67 | 2.00 | 5.97 | 8.00 | 55.00 | 0.25 | 0.50 |
| Horas Remuneradas | 1.00 | 40.00 | 48.00 | 45.18 | 48.00 | 124.00 | 183.99 | 13.56 |
| Horas No Remuneradas | 0.00 | 1.00 | 10.00 | 12.86 | 20.00 | 237.00 | 0.02 | 0.14 |

Como se puede observar en la tabla, mientras la educación y horas remuneradas aún con una gran varianza, aparentan tener la mayoría de sus datos en el centro, y experiencia, tiempo en la empresa y horas no remuneradas con pequeñas varianzas aparentan tener la mayoría de sus datos a la izquierda.

Las siguientes tablas presentan las probabilidades de las variables exógenas del modelo, en estas se puede evidenciar que la gran mayoría de las personas de la muestra no pertenecen a las comunidades indígenas o afro-descendientes, aun si segun el DANE el 4.4 % de la poblacion colombiana se autoreconoce como indígena y 9.34 % para la poblacion autoreconocida como afro, lo que evidencia que estas comunidades no estan bien representadas en desempeño laboral. Adicionalmente solo un 1 % de la población trabajadora es analfabeta. Por otro lado, aproximadamente un 55 % de la población son hombres y un 44 % son mujeres. Por último, tan solo un 2 % aproximadamente de las personas que trabaja están sindicalizadas o tienen un segundo trabajo.

| Probabilidad de las variables exógenas cualitativas | | | | | |
|---|----------|-------------------|----------|---------------|----------|
| Indígena | | Afro-descendiente | | Analfabetismo | |
| 0 | 1 | 0 | 1 | 0 | 1 |
| 0.991742 | 0.008257 | 0.960578 | 0.039421 | 0.988301 | 0.011698 |

| Probabilidad de las variables exógenas cualitativas | | | | | |
|---|----------|-----------------|----------|-----------------|----------|
| Género | | Sindicalización | | Segundo Trabajo | |
| 0 | 1 | 0 | 1 | 0 | 1 |
| 0.553971 | 0.446028 | 0.979551 | 0.020448 | 0.976307 | 0.023692 |

6. Estimación del modelo

6.1. Interpretaciones

- Como era de esperarse y en concordancia con el modelo por cada año de educación el salario aumenta en casi 10 %.
- Por cada año que aumente la experiencia potencial se espera que el salario aumente en un 1,4 %. Por otro lado, en el largo plazo esta variable cambia su tendencia de positiva a negativa, ya que por cada año que aumente la experiencia, el salario presentara una disminución de 0,02 %
- Por cada año que una persona se mantenga trabajando en la misma empresa se espera que su salario aumente en un 1.2 %
- Por cada hora trabajada no remunerada que la persona realice, se tiene que el salario se ve impactado negativamente en un -0.6 %
- Si la persona pertenece a una etnia indígena su salario se verá afectado negativamente en un 8.1 %
- Si la persona pertenece a una etnia afro-descendiente su salario se verá afectado negativamente en un 2.4 %
- Si se es analfabeta el salario se ve aumentado en un 25 %, lo que podría explicarse por una muestra poco representativa, un sesgo en la variable o subjetividad al momento de responder la encuesta.
- El hecho de ser mujer implica que el salario percibido se verá afectado en un 11.7 %
- Si la persona se encuentra sindicalizada, su salario aumenta en un 43.5 % lo que era esperado y es la variable con mayor impacto.
- Contrario a lo que se esperaba, si la persona cuenta con un segundo trabajo, su salario se verá impactado positivamente en un 7 %

6.2. Pruebas de significancia

El modelo es significativo globalmente con un p-value de 0 por lo que tiene significancia a cualquier nivel de confianza, todas las variables cuantitativas son significativas individualmente con un p-value de 0, mientras que dentro de las variables cuantitativas tenemos analfabetismo, género y sindicalización que son significativas con un p-value de 0, segundo trabajo que es significativa con un p-value de 5.2 % por lo que sigue siendo valida trabajando a con un alpha del 10 %. Se continuó estimando el modelo con las variables de autoreconocimiento étnico las cuales no son significativas, no obstante se prefieren mantener en el modelo debido a la relevancia que estas pueden tener para exclusiones sociales por parte de grupos étnicos.

7. Evaluación de supuestos

7.1. Cambio estructural

El cambio estructural hace referencia a un cambio inesperado en una serie de tiempo. Esto puede hacer que se incurra en grandes errores predictivos y poca fiabilidad del modelo en general. Las consecuencias de estimar un modelo por MCO sin tener en cuenta un cambio en la estructura del modelo, representa que los estimadores van a ser sesgados e inconsistentes respecto al comportamiento de cada submuestra; insesgados debido a que el valor de los estimadores tiende a ser diferente del valor real de estos. Otra consecuencia del cambio estructural radica en que el tamaño de los residuos será mayor, esto involucra una SRC mayor, lo que implica un aumento en la varianza y por ende se pierde confiabilidad en el modelo. Al no tener en cuenta el cambio estructural el modelo es susceptible a presentar problemas de heteroscedasticidad o autocorrelación.

Dado que los datos usados para la estimación del modelo son datos de corte transversal, ya que se uso como fuente de información para la estimación del modelo salarial de Micer la Gran Encuesta Integrada de Hogares (GEIH) del Departamento Administrativo Nacional de Estadística (DANE) para diciembre de 2021; ninguna de las pruebas para detectar el cambio estructural resulta pertinente, debido a que por definición este problema solo se puede dar en datos de series de tiempo, es decir, se concluye por definición la no existencia de cambio estructural.

Cuadro 1: Mincer

| | <i>Dependent variable:</i> |
|-------------------------|-----------------------------|
| | ln_ING_H |
| EDUCACION | 0.096*** p = 0.000 |
| EXPERIENCIA | 0.014*** p = 0.000 |
| EXPERIENCIA2 | −0.0002*** p = 0.000 |
| TIEMPO_EMPRESA | 0.012*** p = 0.000 |
| HORAS_REMUNERADAS | −0.009*** p = 0.000 |
| HORAS_NO_REMUNERADAS | −0.006*** p = 0.000 |
| INDIGENA | −0.081 p = 0.222 |
| AFRO | −0.024 p = 0.443 |
| ANALFABETISMO | 0.257*** p = 0.00002 |
| género | −0.117*** p = 0.000 |
| SINDICADO | 0.435*** p = 0.000 |
| SEGUNDO_TRABAJO | 0.077* p = 0.052 |
| Constant | 7.819*** p = 0.000 |
| Observations | 10,172 |
| R ² | 0.318 |
| Adjusted R ² | 0.317 |
| Residual Std. Error | 0.605 (df = 10159) |
| F Statistic | 394.543*** (df = 12; 10159) |
| Breusch-Pagan | 184.83 ($p < 2,2e^{-16}$) |
| Jarque Bera | 68386 ($p < 2,2e^{-16}$) |
| RESET de Ramsey | 153.8 ($p < 2,2e^{-16}$) |
| Durbin-Watson | 1.7489 ($p < 2,2e^{-16}$) |

Note:

*p<0.1; **p<0.05; ***p<0.01

7.2. Correcta especificación

Un error de especificación es el incumplimiento de cualquiera de los supuestos básicos del modelo lineal general. Esta expresión hace referencia al supuesto implícito de que las variables explicativas incluidas en la en el modelo son las verdaderamente relevantes en la explicación de Y . Entre los errores de especificación se encuentra la omisión de variables relevantes o la inclusión de variables irrelevantes. Si se estima el modelo con el primer problema los estimadores tienen a ser sesgados e inconsistentes, mientras que el segundo problema implica estimadores insesgados, pero se pierden grados de libertad injustificadamente. Para detectar si existe un error de especificación se usa el test de RESET (del inglés, Regression Specification Error Test) de Ramsey. Este consiste en realizar una regresión auxiliar del modelo original compuesta por la variables exógenas, más la potencia al cuadrado y al cubo de la variable endógena Y_i , tal que:

$$y_i = \hat{\beta}_1 + \beta_2 \hat{x}_{i2} + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + u_i$$

Donde la hipótesis nula viene dada por $H_0 : \gamma_1 = \gamma_2 = 0$ y representa la no existencia de un error de especificación en el modelo, mientras que la hipótesis alternativa $H_1 : \text{Al menos un } \gamma_i \neq 0 \quad i = 1, 2, \dots$ represente la existencia de un error de especificación. Para el caso del modelo minceriano planteado en este trabajo, el p- value es de $2,2e - 16$, resulto ser inferior a un nivel de significancia de 0,1, debido a lo anterior se rechaza la hipótesis nula, concluyendo la existencia de un error de especificación.

7.2.1. Soluciones al problema de especificación

Para detectar el problema de especificación del modelo se aplicó el test de RESET de forma individual a cada una de las variables exógenas del modelo, la conclusión a las que se llegó fue que las únicas variables que arrojan significancia son las variables cualitativas. Este resultado es un poco ilógico, ya que dentro de las variables cuantitativas se encuentran : los años de educación, los años de experiencia y el tiempo que lleva trabajo la persona en su puesto actual de trabajo, estas variables son relevantes para la estimación del salario en un modelo minceriano.

Las soluciones al error de especificación representa la eliminación de variables irrelevantes, la inclusión de variables relevantes dentro del modelo o el cambio de la forma funcional. Las dos primeras no funcionaron para corregir el cambio, y la última no puede ocasionar el problema, ya que por teoría el modelo de salarios de Mincer, tiene que ser un modelo Log - Lin. Como último recurso para dar solución al problema, se optó por continuar la estimación del modelo con el error de especificación, que en este caso puede no ser tan grave debido a que se está trabajando con un tamaño muestral considerablemente grande, datos de corte transversal y el modelo es globalmente significativo, por tanto el error de especificación no estaría acarreando grandes consecuencias en la interpretación de los parámetros estimados.

7.3. No multicolinealidad

El problema de multicolinealidad consiste en la existencia de relaciones lineales entre dos o más variables independientes del modelo. La multicolinealidad afecta la estimación de los parámetros que acompañan las variables exógenas, cuando existe este problema la matriz de varianzas y covarianzas del modelo no se puede estimar lo que implica una indeterminación de los estimadores.

Para identificar la existencia de un problema de multicolinealidad entre las variables, se debe calcular el índice de correlación entre todas las variables exógenas, dado el caso que alguno tenga un valor superior a 0,9, se concluye un problema de multicolinealidad.

| | | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| afr | -0.01 | -0.03 | 0.01 | 0.01 | 0 | -0.03 | -0.02 | -0.01 | 0.01 | -0.01 | 0 |
| -0.01 | analf | -0.29 | 0.19 | 0.23 | -0.01 | -0.02 | -0.01 | 0.01 | -0.01 | -0.02 | 0.07 |
| -0.03 | -0.29 | educ | -0.53 | -0.55 | 0.02 | 0.02 | -0.04 | 0.02 | 0.1 | 0.11 | -0.12 |
| 0.01 | 0.19 | -0.53 | exp | 0.95 | -0.02 | -0.07 | 0.01 | 0 | -0.06 | 0.04 | 0.47 |
| 0.01 | 0.23 | -0.55 | 0.95 | exp2 | -0.04 | -0.09 | 0.01 | 0 | -0.06 | 0.02 | 0.45 |
| 0 | -0.01 | 0.02 | -0.02 | -0.04 | hnr | -0.27 | 0 | 0.04 | 0.5 | -0.02 | -0.07 |
| -0.03 | -0.02 | 0.02 | -0.07 | -0.09 | -0.27 | hr | 0 | 0.05 | -0.19 | 0.01 | 0.03 |
| -0.02 | -0.01 | -0.04 | 0.01 | 0.01 | 0 | 0 | ind | -0.01 | -0.01 | -0.01 | -0.02 |
| -0.01 | 0.01 | 0.02 | 0 | 0 | 0.04 | 0.05 | -0.01 | segt | 0.04 | 0.02 | -0.02 |
| 0.01 | -0.01 | 0.1 | -0.06 | -0.06 | 0.5 | -0.19 | -0.01 | 0.04 | sexo | 0 | -0.11 |
| -0.01 | -0.02 | 0.11 | 0.04 | 0.02 | -0.02 | 0.01 | -0.01 | 0.02 | 0 | sind | 0.15 |
| 0 | 0.07 | -0.12 | 0.47 | 0.45 | -0.07 | 0.03 | -0.02 | -0.02 | -0.11 | 0.15 | tmpe |

Figura 1: Matriz de correlaciones de variables endógenas

De acuerdo con la gráfica anterior se puede afirmar que para el modelo de salarios de tipo minceriano que se presenta en este documento, no hay problemas de multicolinealidad, ya que las únicas correlaciones que dan un valor superior a 0,9 son las de las variables experiencia y experiencia al cuadrado.

7.4. Homoscedasticidad

El supuesto de homoscedasticidad implica que, condicionando en las variables explicativas, la varianza del término de error no observado es constante, la violación a este supuesto es la heteroscedasticidad, donde la varianza del error es diferente para cada valor de x . Estimar un modelo con heteroscedasticidad implica tener estimadores MCO insesgados, además, como la $Var(u|X)$ ya no es constante, el estimador MCO ya no es MELI (Mejores Estimadores Lineales Insesgados) y el estimador MCO ya no es asintóticamente eficiente.

Para detectar un problema de heteroscedasticidad, está la prueba propuesta por Breusch and Pagan (1979), que consiste en ajustar un modelo de regresión lineal con variables respuesta dadas por residuales del modelo original al cuadrado

$$\hat{u}_i^2 = \gamma_1 + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \dots + \gamma_k x_{ik}$$

y como covariables las variables del modelo original. Si se concluye que $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_k = 0$, significa que los residuales no son función de las covariables del modelo. El estadístico en esta prueba está dado por $n * R^2$ y bajo la hipótesis nula verdadera, el estadístico tiene distribución χ_k^2 .

Para el caso del modelo tratado en este documento, dado que el p-value ($2,2e-16$) es menor que un nivel de significancia de 0,1 se concluye que existe heteroscedasticidad.

Para corregir los problemas sobre los estimadores derivados de la heteroscedasticidad se utiliza el método de mínimos cuadrados generalizados o ponderados.

7.5. No autocorrelación

La autocorrelación se define como la perturbación de una observación cualquiera u_i que está correlacionada con la perturbación de cualquier otra observación, es decir que las observaciones no son independientes. La estimación del modelo con violación a este supuesto implica fallas en los estimadores por MCO, puesto que estos ya no serán insesgados y de varianza mínima, por otra parte, la inferencia estadística que se obtiene de la matriz de varianzas y covarianzas no es correcta.

Para detectar este problema se usa la prueba Durbin - Watson, que detecta autocorrelación de primer orden : para el caso del modelo de salarios minceriano planteado durante el presente trabajo, la prueba realizada por R utilizando este test indica que existe un problema de autocorrelación espacial (datos de corte transversal) con una significancia del 0,1 , puesto que el p-valor es de $2,2e - 16$, sin embargo el problema de autocorrelación no es tan grave, ya que el estadístico Durbin - Watson es de 1,7489, está orientado a concluir no autocorrelación de primer orden, debido a que relativamente cercano a 2, además la violación a este supuesto deber estar siendo ocasionada por la inclusión de la variable exp^2 por definición teórica del modelo minceriano. Es decir que a pesar de que existe autocorrelación es comúnmente válida para este tipo de modelos y datos.

7.6. Normalidad

Uno de los supuestos básicos del modelo de regresión lineal clásico es el que los errores tengan distribución normal, si este supuesto no se cumple, no se tiene la justificación teórica para la utilización de pruebas estadísticas que involucren a las distribuciones t , F y χ^2 .

La prueba Jarque-Bera (1987) permite evidenciar el supuesto de normalidad, consiste en obtener un valor estadístico que será contrastado con una distribución χ^2 con dos grados de libertad, con base en la asimetría y la curtosis o apuntamiento de la gráfica de distribución de los residuos MCO. Al realizar esta prueba al modelo tratado en el presente texto, se encontró un p-valor de $2.2e-16$, por lo que se concluye no normalidad para el modelo.

7.7. Intento de solución de Normalidad, heteroscedasticidad y no especificación

Sin ningún fruto se intentó solucionar los errores de Normalidad, heteroscedasticidad y no especificación, no obstante ante cada intento el p-value de cada uno de estos métodos continuaba en 0, por lo que resulto ineficiente el planteamiento de estos resultados.

Para intentar solucionar el problema de no normalidad de los errores, se intentó primero eliminar todos los datos atípicos de la variable exógena así como todos los datos con niveles de errores atípicos no obstante esto no logro siquiera aumentar el p-value que se mantuvo en 0.

Para intentar solucionar los problemas de heteroscedasticidad y no especificación se intentó realizando tanto regresión mediante mínimos cuadrados generalizados como mínimos cuadrados ponderados, en ninguno de los dos casos se obtuvo un resultado gratificante y por el contrario todos los problemas se mantenían.

Como conjetura se cree que el problema puede ser ocasionado primeramente debido a que el modelo minceriano tiene la variable de experiencia² lo que puede ocasionar que no se logre una buena regresión mediante MCO, otra posibilidad es que estos errores sean debido a la subjetividad e inexactitud de las encuestas que nos puede llevar a obtener datos fuertemente sesgados por algunos, ya sea afirmando tener mayores o menores salarios o ocultando datos relevantes sobre sus características lo que puede llevarnos a cometer fuertes errores al momento de realizar la regresión.

Dado que los cambios realizados no lograron solucionar los problemas de violación de los supuestos, la estimación de los parámetros del modelo e interpretación de los mismos.

8. Predicción y ajuste

8.1. Predicción

Debido a la no fiabilidad del modelo por lo anteriormente expuesto se decidió realizar una simulación de Montecarlo para evaluar si al menos la regresión realizada es capaz de aproximarse a la media de la muestra, con una simulación de 10mil individuos se obtuvo la siguiente distribución.

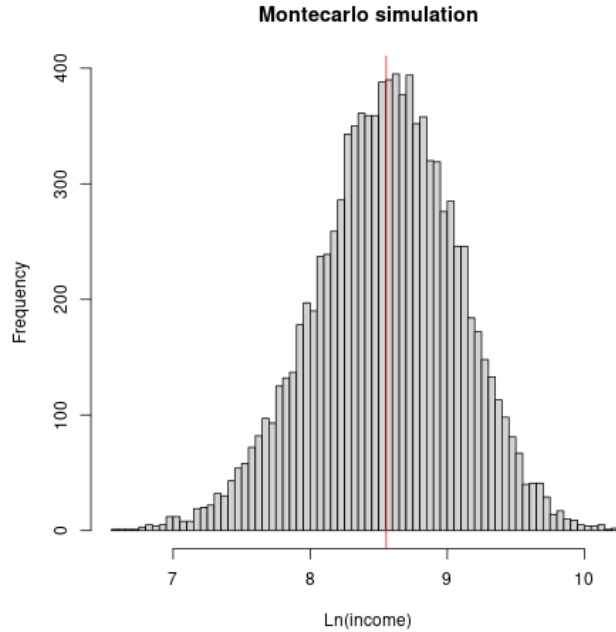


Figura 2: distribución del ingreso simulado

Con sorpresa la simulación logro el cometido y obtuvo una media de 8.55 cuando la media de la muestra real es 8.56 por lo que logro aproximarse de forma cuasiperfecta aún con todos los problemas que presenta el modelo.

8.2. Ajuste

Debido a que los datos utilizados son de corte transversal no se puede realizar un ajuste de los datos, ya que este ajuste es funcional para datos de series de tiempo o sus derivados tales como datos panel.

9. Conclusiones

Al finalizar el modelo se puede concluir que las variables exógenas, Educación, experiencia, tiempo que lleva trabajo en la empresa, sindicalización y segundo trabajo, afectan de manera positiva el salario, mientras que la variable horas no remuneradas y las variables cualitativas autorreconocerse como miembro de comunidad indígena o afro-descendiente y ser mujer, afecta de forma negativa el salario, estos resultados fueron obtenidos mediante el modelo sin tener en cuenta las violaciones a los supuestos, por las razones anteriormente explicadas.

El único parámetro poco lógico que resalta en la estimación es el analfabetismo, la interpretación de este parámetro indica que si una persona es analfabeta su salario podría aumentar en un 25 %, situación que claramente no concuerda con la realidad, ni con el Modelo de salarios de Mincer, debido a que en este modelo se asume que ser analfabeta es no tener nivel de educación. Esta anomalía se puede explicar por una muestra poco representativa o a un sesgo en la variable, debido a la subjetividad al momento de contestar la encuesta.

Por otro lado, se puede concluir que debido a que los datos son de corte transversal, los supuestos de autocorrelación, heteroscedasticidad y no normalidad, puede que no representen grandes problemas para obtener una visión simplificada de las variables que impactan en los salarios y de las interpretaciones de las mismas, donde la mayoría concuerdan con el modelo minceriano. Esto va respaldado con el hecho de que el modelo es global e individualmente significativo, a excepción de la variables cualitativas autoreconocimiento como indígena o afro-descendiente.

10. Referencias

Vela Peón, F. (2010). Normalidad de los errores. Universidad Autónoma Metropolitana.
<https://mregresion.files.wordpress.com/2011/10/normalidad.pdf>

Miranda, J. (2017, 8 agosto). Autocorrelación. Todo econometría.
<https://todoeconometria.wordpress.com/2017/08/08/autocorrelacion/>

Hernández, F. (2020, 30 octubre). 11 Pruebas de Homocedasticidad — Modelos de Regresión con R. Modelos de regresión en R. <https://fhernanb.github.io/libroregresion/homo.html>

Moreno, P., Rodriguez, J. M., Soberon, A. (s. f.). Econometría I — Tema 6: Heterocedasticidad. Econometría I.
<https://ocw.unican.es/pluginfile.php/1127/course/section/1352>

Salmerón, R. (s. f.). Multicolinealidad. Multicolinealidad, Universidad de Granada.
<https://www.ugr.es/~romansg/material/WebEco/02-Eco/Teoria/tema4.pdf>

Bour, Enrique A. (Sf). La Ecuación de J. Mincer.
Recuperado de: [http://ebour.com.ar/mec_abogados/Bour %20- %20La %20ecuacion %20de %20Mincer.pdf](http://ebour.com.ar/mec_abogados/Bour%20-%20La%20ecuacion%20de%20Mincer.pdf)

Freire y Tejeiro. (Sf). Las ecuaciones de Mincer y las tasas de rendimiento de la educación en Galicia.
Recuperado de: <https://2010.economicsofeducation.com/user/pdfsiones/095.pdf>