

ANÁLISIS MULTICAUSAL DE LOS
DETERMINANTES DEL INGRESO
LABORAL EN BOLIVIA PARA EL 2023.
UN ENFOQUE CON MODELOS DE VARIABLES
INSTRUMENTALES BIETAPICO

*MULTICAUSAL ANALYSIS OF THE
DETERMINANTS OF LABOR INCOME IN BOLIVIA
FOR 2023, A TWO-STAGE INSTRUMENTAL
VARIABLE MODELS APPROACH*

Benigno Caballero Claire¹
Rolando Caballero Martínez²

-
- 1 Universidad Técnica de Oruro, University of Albuquerque (EE.UU.). Miembro de la Red Académica de América Latina y el Caribe sobre China (RED ALC-CHINA)
- 2 Universidad Nacional Autónoma de México. Centro de Estudios Monetarios y Financieros (CEMFI, Madrid, España). Universidad Católica Boliviana (Santa Cruz, Bolivia). Universidad Técnica de Oruro. Universidad Autónoma Gabriel René Moreno. Escuela Militar de

RESUMEN

Este documento analiza los posibles factores que podrían influir como determinantes principales de los ingresos laborales de Bolivia en el periodo 2023 y para el mismo se utiliza la base de datos de la encuesta de hogares MECOVI 2023 y presenta evidencia que, en Bolivia durante 2023, los ingresos laborales estuvieron determinados principalmente por el nivel educativo, la edad, el género y el área geográfica de donde proviene el encuestado. Estos factores marcaron diferencias significativas en salarios y oportunidades, reflejando desigualdades persistentes en el mercado laboral, para tal efecto se utiliza inicialmente modelos de regresión lineal mínimos cuadrados ordinarios (MCO), para finalmente utilizar modelos de regresión con variables instrumentales en dos etapas con el objetivo de corregir posible presencia de endogeneidad. Los resultados demuestran, que los diferentes factores mencionados anteriormente ayudan a explicar e diferente grado de significancia la estructura de los ingresos laborales en Bolivia.

Palabras clave: Salarios, Empleo y Productividad, Trabajo.

Clasificación JEL: J31, J23, J20

ABSTRACT

This document analyzes the potential factors that could influence labor income in Bolivia during 2023. Using the MECOVI 2023 household survey database, it presents evidence that, in Bolivia during 2023, labor income was primarily determined by educational level, age, gender, and geographic origin. These factors resulted in significant differences in wages and opportunities, reflecting persistent inequalities in the labor market. Initially, ordinary least squares (OLS) linear regression models were used, followed by two-stage instrumental variables regression models to correct for potential endogeneity. The results demonstrate that the aforementioned factors contribute to explaining the structure of labor income in Bolivia to varying degrees.

Keywords: Wages, Employment and Productivity, Labor.

JEL Classification: J31, J23, J20

1. Introducción

En Bolivia, los principales determinantes del ingreso laboral son el nivel educativo, la edad, el género, la experiencia laboral y la estructura del empleo. Estos factores influyen directamente en la desigualdad salarial y en el crecimiento económico del país.

En las últimas décadas, el área conurbada de Bolivia ha presentado mejoras significativas en la contracción de la pobreza. Pese a este aspecto, estos avances no se han reflejado en mayores ingresos ni menores niveles de desigualdad salarial (ver Muriel y Jemio, 2009).

En este contexto, surge la necesidad de analizar con mayor profundidad los factores que explican los ingresos laborales. Diversos estudios han sido desarrollados para conocer los factores que explican los ingresos laborales urbanos en Bolivia; tomando en cuenta las llamadas variables de oferta relacionadas con la teoría del capital humano — educación y experiencia — así como los problemas de discriminación y segregación por género y etnia (ver Pérez de Rada, 1997; Fields et al, 1998; Rivero y Jiménez, 1999; Moensted, 2000).

En muchos casos se apunta a los años de escolaridad como la variable más importante; ya que explica la mayor parte de la desigualdad salarial

estimada (Fields et al. 1998; Spatz y Steiner, 2002; y Andersen, 2001).

Principales determinantes del ingreso laboral en Bolivia

1. Nivel educativo

- El capital humano es el factor más influyente: a mayor nivel de educación, mayor ingreso laboral.
- Los trabajadores con educación universitaria o técnica suelen percibir salarios significativamente más altos que aquellos con solo educación primaria o secundaria.

2. Edad y experiencia

- La edad está asociada con la experiencia acumulada en el mercado laboral.
- Los ingresos tienden a aumentar con la edad hasta cierto punto, pero pueden estancarse o disminuir en etapas avanzadas de la vida laboral.

3. Género y brecha salarial

- Existe una brecha salarial de género persistente: las mujeres, en promedio, ganan menos que los hombres por trabajos similares.
- Factores culturales y estructurales del mercado laboral refuerzan esta desigualdad.



4. Estructura del empleo

- La informalidad laboral en Bolivia es alta, lo que limita el acceso a salarios estables y beneficios sociales.
- Los trabajadores en sectores formales (industria, servicios profesionales) tienen ingresos superiores a los del sector informal (comercio minorista, agricultura).

5. Región y contexto socioeconómico

- Los ingresos varían según la región: en ciudades como La Paz, Santa Cruz y Cochabamba los salarios tienden a ser más altos que en áreas rurales.
- Factores como el acceso a infraestructura, mercados y oportunidades de empleo influyen en estas diferencias.

Impacto en el crecimiento económico

- Los determinantes del ingreso laboral están directamente vinculados al crecimiento económico nacional.
- Una mayor inversión en educación y reducción de la brecha de género pueden mejorar la productividad y reducir la desigualdad.
- La informalidad laboral, en cambio, limita la capacidad del Estado para recaudar impuestos y financiar políticas públicas.

El ingreso laboral en Bolivia depende de múltiples factores, pero los más decisivos son la educación, la experiencia y el género. Reducir la informalidad y promover la igualdad de oportunidades son claves para mejorar los ingresos y fortalecer el desarrollo económico del país.

2. Metodología y Análisis Estadístico Descriptivo

El estudio o el análisis de los ingresos laborales a partir de la Encuesta de Hogares 2023 en Bolivia se considera importante porque ayuda a conocer la situación económica de la población a partir de otras variables de interés que influyen en la variable objeto de estudio, ayuda a ver o identificar diferencias entre grupos sociales, por género, analizar otros factores como la experiencia o lugar de residencia por ejemplo departamento. Su análisis ayuda a entender de mejor manera las fuentes o diferencias de los ingresos laborales existentes en Bolivia.

De acuerdo a la literatura económica, la experiencia laboral es una variable clave o determinante para explicar el ingreso laboral ya que muestra o refleja la acumulación de habilidades, conocimientos y productividad logrados a lo largo del tiempo. En el caso boliviano, en la que una parte importante de la fuerza laboral accede al mercado de trabajo

sin formación profesional avanzada, entonces la experiencia se convierte en una variable decisiva para aumentar los ingresos, consolidar estabilidad laboral y diferenciar niveles salariales a ser analizadas en el presente trabajo.

Al no existir explícitamente la variable experiencia laboral en la base de datos de encuesta de hogares 2023, se procede a calcular la variable para cada observación de la siguiente manera:

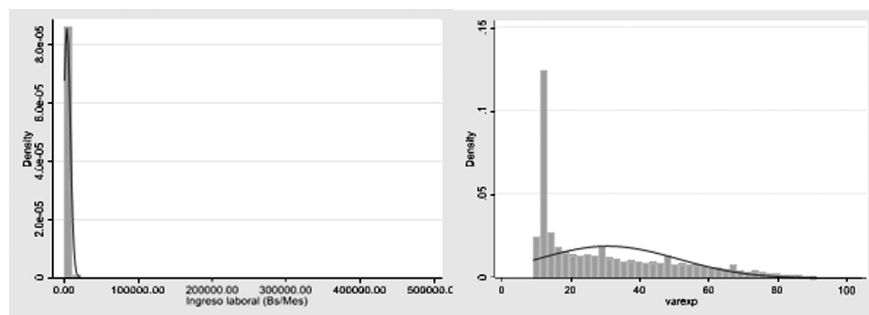
$$\text{varexp} = \text{s01a_03} - (\text{aestudio} - 6) \text{ Experiencia Laboral} = \text{Años cumplidos} - \text{Años de estudio} - 6$$

Análisis descriptivo básico de las variables de estudio

Variable	Obs	Mean	Std. Dev.	Min	Max
ylab	16,65	3169,42	4642,41	5	480422,2
carexp	38,37	30,50	20,31	9	104

El ingreso laboral (Ylab) muestra una dispersión grande ya que la desviación estándar es mayor que la media, además tiene distribución sesgada hacia la derecha y la media esta influenciada por los valores extremadamente altos, entonces la variable presenta una gran variabilidad.

Por su parte la variable experiencia (varexp) muestra menor dispersión, al parecer la media es más representativa, en resumen, es una variable más estable que el ingreso laboral, lo que muestra en los siguientes gráficos.



Primera regresión: ylab y experiencia

ylab	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
varexp	-14.523	2.149	-6.76	0	-18.735	-10.311	***
Constant	3.706.412	87.204	42.50	0	3.535.483	3877.34	***
Mean dependent var	3.169.422		SD dependent var	4.642.413			
R-squared	0.0027		Number of obs	16653			
F-test	45.671		Prob > F	0.000			
Akaike crit. (AIC)	328.41 8.760		Bayesian crit. (BIC)	328.434.200			
*** $p < .01$, ** $p < .05$, * $p < .1$							

Análisis de resultados

Existe una relación negativa inesperada o coeficiente de varexp -14,52, como las variables están en nivel-nivel, un año adicional de experiencia laboral disminuye el ingreso laboral en 14,52 bolivianos. Este resultado contradice la teoría económica y se debe buscar posibles causas, este hecho puede ocurrir por los siguientes motivos: variables omitidas, problemas de medición de la experiencia, entre otros. Cuando se ve la calidad del modelo en función de los resultados, R², solo el 0,27% de la variación del ingreso laboral es

explicada por la experiencia laboral, entonces es necesario insertar otras variables que ayuden a explicar mejor el ingreso laboral.

3. Resultados Econométricos

Para mejorar el modelo, se selecciona otras variables de interés, que en función de la teoría económica sí podrían afectar o estar relacionadas con las variables de interés que es el ingreso laboral, por lo que se seleccionan las siguientes variables: años de estudio (aestudio), área (urbano o rural), género (hombre o mujer, s01a_02).

$$ylab_i = \theta_0 + \theta_1 experiencia_i + \theta_2 estudio_i + \theta_3 area_i + \theta_4 genero_i + \epsilon_i$$

.reg ylab varexp estudio i.area i.s01a_02

Source	SS	df	MS	Number of obs	=	16,653
Model	2.0433e+10	4	5.1083e+09	F (4,16648)	=	251.27
Residual	3.3845e+11	16,648	20329799	Prob > F	=	0.0000
				R-squared	=	0.0569
Total	3.5888e+11	16,652	21551994.2	Adj R-squared	=	0.0567
				Root MSE	=	4508.9

ylab	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
varexp	30.70678	2.622609	11.71	0.000	25.56619	35.84737
aestudio	226.9212	9.211241	24.64	0.000	208.8662	244.9762
area Rural	-713.8639	91.4386	-7.81	0.000	-893.0933	-534.6345
s01a_02	-801.3219	71.73739	-11.17	0.000	-941.9348	-660.709
2. Mujer_cons	-9.511674	186.7418	-0.05	0.959	-375.5455	356.5221

Justificación teórica de la selección de las variables de control con estadísticos pertinentes (significancia, R2 ajustado, etc.)

- **Años de estudio (aestudio)**, fundamentada en la teoría del capital humano (Becker, Mincer). A mayor educación, mayor productividad y por tanto se espera mayores ingresos.

literatura económica indica la diferencia salarial de género.

A continuación, se realiza la regresión del modelo:

- **Experiencia laboral (varexp)**, de igual manera que el anterior, la experiencia permite adquirir habilidades prácticas que ayudan a incrementar la productividad, lo que puede ayudar a mejorar los ingresos.

Se observa que todas las variables son significativas al 1% y al 5%, asimismo **F (4, 16648) = 251.27**, Prob > F = 0.0000 muestra que el modelo en su conjunto es estadísticamente significativo. El **R-cuadrado = 0.0569 (5.7%)** muestra que el modelo explica una fracción pequeña de la variabilidad de la variable dependiente (**ylab**). Esto sugiere que hay muchos otros factores no incluidos que afectan al resultado. Para mejorar los resultados obtenidos, se realizan ajustes a las variables experiencia, años de estudio e ingreso laboral.

- **Área de residencia (urbano/rural)**, diferencias estructurales en el trabajo; en zonas rurales de Bolivia predominan actividades agrícolas de baja remuneración, mientras que en áreas urbanas hay mayor acceso a empleos formales y normalmente mejor remuneradas.

- **Sexo (mujer/hombre)**, ver por diferencias de género en ingresos es fundamental dado que la

De acuerdo a la regresión realizada en el anterior inciso, se obtiene lo siguiente:

$$y_{lab_i} = -9.5117 + 30,71 \text{ experiencia}_i + 226.92 \text{ aestudio}_i - 713,86 \text{ area (rural)}_i - 801,32 \text{ genero (mujer)}_i + \varepsilon_i$$

Asimismo, el análisis de los coeficientes de la regresión se muestra en la siguiente tabla:

Variable	Coef.	Interpretación
varexp (experiencia)	30.71	Cada año adicional de experiencia laboral aumenta en promedio el ingreso laboral en aproximadamente 30.7 bolivianos (ceteris páribus). Es significativo ($p < 0.001$).
aestudio (años de estudio)	226.92	Cada año adicional de educación incrementa el ingreso laboral en aproximadamente 227 bolivianos . Significativo ($p < 0.001$).
Rural (vs Urbano)	-713.86	Una persona que vive en área rural solo por pertenecer a dicha área en promedio gana aproximadamente menos 714 bolivianos respecto a alguien del área urbana con las mismas características. Significativo ($p < 0.001$).
Mujer (vs Hombre)	-801.32	Una mujer gana en promedio menos 801 bolivianos que un hombre con las mismas características. Significativo ($p < 0.001$).
Constante (_cons)	-9.51	Representa el ingreso esperado cuando todas las variables explicativas son cero (poco interpretable en este caso).

Como se observó en el anterior inciso, el R cuadrado era muy bajo por lo que se optó por añadir logaritmos a las variables que tenían mayor dispersión, en este caso: ylab, experiencia y años de estudio. Es necesario logaritmizar las variables porque así se puede:

1) Reducir asimetría: La variable ingresos de acuerdo al gráfico anterior se encuentra muy **sesgada** (muchos trabajadores con salarios bajos y pocos con salarios muy altos). El logaritmo "estrecha" los valores extremos y produce una distribución más cercana a la normal, lo que mejora la validez de los supuestos de MCO.

2) Reducir heterocedasticidad: La varianza de los ingresos tiende a crecer con el nivel de ingreso (los ricos tienen más dispersión salarial).

Por lo tanto, tomar logaritmos ayuda a estabilizar la varianza, reduciendo problemas de heterocedasticidad. Luego de generar los logaritmos correspondientes a las variables, corremos el siguiente modelo:

$$\log(y_{lab}_i) = \theta_0 + \theta_1 \log(experiencia_i) + \theta_2 \log(aestudio_i) + \theta_3 \text{area}(rural)_i + \theta_4 \text{genero}(mujer)_i + \varepsilon_i$$

```
.reg logylab logexp logaest i.area i.s01a_02
```

Source	SS	df	MS	Number of obs	=	16,200
Model	2809.02934	4	702.257336	F(4,16195)	=	1173.15
Residual	9694.48017	16,195	.598609458	Prob > F	=	0.0000
				R-squared	=	0.2247
				Adj R-squared	=	0.2245
Total	12503.5095	16,199	.771869221	Root MSE	=	.7737

logylab	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
logexp	.2123828	.0154879	13.71	0.000	.1820249	.2427408
logaest	.5475074	.0128321	42.67	0.000	.5223551	.5726598
area Rural	-.5229903	.0160383	-32.61	0.000	-.5544272	-.4915534
s01a_02	-.271042	.0125492	-21.60	0.000	-.2956399	-.2464441
_cons	5.962813	.0744983	80.04	0.000	5.816788	6.108838

4. Análisis general del modelo

$$\log(y_{lab}_i) = 5,96 + 0,21 \log(experiencia_i) + 0,54 \log(aestudio_i) - 0,52 \text{area}(rural)_i - 0,27 \text{genero}(mujer)_i + \varepsilon_i$$

- (**R²**): 0.2247, el modelo explica aproximadamente el 22,5% de la variabilidad de logylab.

- **Adj R-squared:** 0.2245, parecido a **R²**, ajustado por el número de variables.
- **F(4,16195) = 1173.15, p < 0.001**
→ El modelo es globalmente significativo; lo que significa que al menos una variable independiente tiene efecto sobre logylab.

Variable	Coefficiente	Interpretación
logexp	0,212	Un aumento del 1% en la experiencia laboral genera un aumento aproximado del 0.21% en los ingresos laborales (logylab), manteniendo las demás variables constantes.
logaest	0,547	Un aumento del 1% en los años de estudio está asociado con un aumento del 0.547% en los ingresos laborales. Esto indica que la educación tiene un efecto más fuerte de acuerdo a los resultados del modelo.
area (Rural)	-0,523	Vivir en área rural implica que, en promedio, el ingreso laboral es menor en $(e^{-0,523} - 1) \times 100 \approx -40,7\%$ respecto a un urbano.
s01a_02 (Mujer)	-0,271	Ser mujer está asociado con un ingreso laboral $(e^{-0,271} - 1) \times 100 \approx -23,8\%$ menor que el de los hombres, manteniendo todo lo demás constante.
_cons	5,963	Este es el valor esperado de logylab (ingreso laboral) cuando todas las variables independientes son 0.



Además, todas las variables tienen $p < 0.001$, por lo que son estadísticamente significativas al 1%, es decir el modelo es significativo de manera conjunta como de manera individual. En cada caso los intervalos de confianza [95% conf. interval] no incluyen el cero, lo que confirma la significancia de los coeficientes, indicado arriba.

• Test De Heterocedasticidad Breusch Pagan/Cook–Weisberg

```
. estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of logylab

H0: Constant variance

chi2(1) = 1475.49

Prob > chi2 = 0.0000
```

El test de Breusch–Pagan/Cook–Weisberg arroja un estadístico $\chi^2(1)=1475.49$ con $p=0.0000$, lo que implica que se rechaza la hipótesis nula de homocedasticidad y, por tanto, el modelo presenta heterocedasticidad. Esto significa que la varianza de los errores no es constante, lo que no sesga los coeficientes estimados pero sí vuelve poco confiables sus errores estándar y, en consecuencia, las pruebas de significancia. Para corregir este problema, es recomendable utilizar estimaciones con errores robustos.

White Test

```
. estat imtest, white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

chi2(12) = 1588.21
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	P
Heteroskedasticity	1588.21	12	0.0000
Skewness	277.47	4	0.0000
Kurtosis	99.02	1	0.0000
Total	1964.70	17	0.0000

El test de White confirma la presencia de heterocedasticidad en el modelo, con un estadístico $\chi^2=1588.21$ y $p=0.0000$, lo que lleva a rechazar la hipótesis nula de homocedasticidad. La descomposición de Cameron y Trivedi refuerza este resultado al mostrar además que los errores presentan asimetría y exceso de curtosis, lo que indica que no siguen una distribución normal. En conjunto, estos hallazgos implican que, si bien los coeficientes estimados permanecen insesgados, los errores estándar y las pruebas de significancia no son confiables bajo los supuestos clásicos, por lo que se recomienda emplear errores robustos u otros métodos que corrijan la heterocedasticidad y la no normalidad.

• Multicolinealidad

```
. vif
```

Variable	VIF	1/VIF
logexp	1.38	0.723390
logaest	1.51	0.661157
2.area	1.15	0.871744
2.s01a_02	1.02	0.984937
Mean VIF	1.26	

Los valores de VIF (Factor de Inflación de la Varianza) muestran que no hay problemas significativos de multicolinealidad entre las variables incluidas en el modelo. Todos los VIF individuales son bajos, con logaest siendo el más alto en 1.51, lo que indica que la varianza de sus coeficientes no se ve inflada de manera preocupante por la correlación con otras variables. El valor medio de VIF de 1.26 confirma que, en general, las variables explicativas están suficientemente independientes entre sí, por lo que los estimadores de regresión son confiables y la multicolinealidad no representa un riesgo para la interpretación de los resultados.

• Prueba de Normalidad

```
. swilk resid
```

Shapiro-wilk W test for normal data

Variable	Obs	W	V	Z	Prob>=z
resid	16,200	0.98018	148.796	13.557	0.00000

Note: The normal approximation to the sampling distribution of W is valid for $4 < n \leq 2000$.

```
. sktest resid
```

Skewness and kurtosis tests for normality

—Joint test—

Variable	Obs	Pr(skewness)	Pr(kurtosis)	Adj chi2(2)	Prob>chi2
resid	16,200	0.0000	0.0000	844.72	0.0000

Los residuos no cumplen el supuesto de normalidad, lo cual podría afectar pruebas de hipótesis basadas en t y F si la muestra fuera pequeña. Sin embargo, dado que tu muestra es muy grande ($n = 16,200$), el incumplimiento de normalidad es menos preocupante,

ya que los estimadores OLS siguen siendo consistentes y las pruebas de significancia se aproximan bien por el Teorema Central del Límite.

El modelo presenta heterocedasticidad, por lo tanto, se emplean errores robustos para corregir este problema.

La regresión robusta de ingresos laborales sobre experiencia laboral, años de estudio, área y sexo muestra que todos los coeficientes son estadísticamente significativos: la educación y la experiencia aumentan los ingresos logarítmicos, mientras que vivir en área rural o ser mujer los reduce. El modelo explica alrededor del 22,5% de la variabilidad en los ingresos y, al usar errores estándar robustos, las inferencias son confiables incluso ante heterocedasticidad o residuos no normales. En general, los resultados confirman la relevancia de la educación, la experiencia, el área y el género como determinantes significativos del ingreso laboral.

Utilizar una variable instrumental para corregir posibles errores por endogeneidad.

Se procede a añadir la variable instrumental para experiencia, en este caso se escoge "Edad inicio de trabajo", calculada como la edad menos la experiencia:

$$\text{Edad inicio de trabajo} = \text{Edad} - \text{Experiencia}$$



```
. reg logylab logexp logaest 1.area 1.s01a_02, robust
```

Linear regression

Number of obs = 16,200

F(4, 16195) = 809.81

Prob > F = 0.0000

R-squared = 0.2247

Root MSE = .7737

logylab	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
Logexp	.2123828	.0164062	12.95	0.000	.1802249 .2445407
Logaest	.5475074	.0153427	35.69	0.000	.5174341 .5775808
area Rural	-.5229903	.0198243	-26.38	0.000	-.5618482 -.4841324
s01a_02					
2.Mujer	-.271042	.0124273	-21.81	0.000	-.2954009 -.2466831
_cons	5.962813	.0817689	72.92	0.000	5.802537 6.123089

Se tomó la variable “Edad inicio de trabajo” debido a las siguientes razones:

- **Relevancia:** la edad de inicio laboral está fuertemente correlacionada con la experiencia acumulada, lo cual se confirma tanto teóricamente (capital humano de Becker y Mincer).
- **Exogeneidad:** no afecta de manera directa el ingreso actual más allá de la experiencia, pues su único canal de influencia es la acumulación de años trabajados. Además, al controlar por educación, género y área de residencia, se refuerza el cumplimiento de la **restricción de exclusión**.

Asimismo, para tener una menor variación de los valores de esta variable, se aplicaron logaritmos.

Primera Etapa

```
. reg logylab logexp logaest 1.area 1.s01a_02, robust
```

```
. **ESTIMACION CON VARIABLES INSTRUMENTALES
```

```
. *PRIMERA ETAPA:
```

```
. ivregress 2sls logylab (logexp = logedad_inicio) logaest 1.area 1.s01a_02, first
```

First-stage regressions

Number of obs = 12,906

F(4, 12901) = 132.44

Prob > F = 0.0000

R-squared = 0.0394

Adj R-squared = 0.0391

Root MSE = 0.4107

logexp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
logaest	.3446908	.0691	4.99	0.000	.2092447 .480137
area Rural	-.0279673	.0104991	-2.66	0.000	-.048547 -.0073875
s01a_02					
2.Mujer	-.0322658	.0074953	-4.30	0.000	-.0469578 -.0175739
logedad_inicio	-.290995	.0303238	-9.60	0.000	-.3504341 -.231556
_cons	3.040256	.1222748	24.86	0.000	2.80058 3.279933

$$\ln(\exp_i) = 3,040 + (-0,291)\ln(\text{edad}_{\text{inicio}_i}) + 0,345\ln(\text{aest}_i) - 0,028\text{Rural}_i - 0,032\text{Mujer}_i + u_i$$

En esta etapa, la variable endógena **experiencia laboral** ($\ln(\exp)$) se regresa sobre el instrumento **edad de inicio laboral** ($\ln(\text{edad_inicio})$) y las variables exógenas (**años de educación, área y género**). El coeficiente negativo de $\ln(\text{edad_inicio})$ indica que quienes comienzan a trabajar más tarde acumulan menos experiencia, confirmando la relevancia del instrumento. Los demás controles muestran efectos menores pero significativos: mayor educación aumenta la experiencia, mientras que residir en área rural o ser mujer la reduce ligeramente.

El instrumento **logedad_inicio** es confiable. En la primera etapa, su coeficiente es **-0.291** con un **t = -9.60** y **P = 0.000**, lo que indica que está fuertemente correlacionado con la experiencia y cumple la condición de relevancia; además, el **F estadístico global es 132.44**, muy superior al

umbral de 10, confirmando que no es un instrumento débil. Teóricamente, la edad de inicio laboral afecta la experiencia acumulada pero no influye directamente en los ingresos actuales más allá de ese efecto, y al controlar por educación, área y género, la exclusión parece plausible. Por estas razones, **logedad_inicio es un instrumento fuerte y válido para estimar el efecto causal de la experiencia sobre los ingresos**.

El resultado de esta etapa es la **experiencia instrumentada** ($\hat{\ln}(\exp)$), que se usará en la segunda etapa.

Segunda etapa (2SLS, regresión estructural)

$$\ln(Y_{lab}_i) = -5,849 + 2,718\hat{\ln}(\exp_i) + 1,842\ln(\text{aest}_i) - 0,197\text{Rural}_i - 0,234\text{Mujer}_i + u_i$$

```
. reg logylab logexp logaest 1.area 1.s01a_02, robust
. **ESTIMACION CON VARIABLES INSTRUMENTALES
. *PRIMERA ETAPA:
. ivregress 2sls logylab (logexp = logedad_inicio) logaest 1.area 1.s01a_02, first
```

First-stage regressions

Number of obs = 12,906
 F(4, 12901) = 132.44
 Prob > F = 0.0000
 R-squared = 0.0394
 Adj R-squared = 0.0391
 Root MSE = 0.4107

logexp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
logaest	.3446908	.0691	4.99	0.000	.2092447 .480137
area Rural	-.0279673	.0104991	-2.66	0.000	-.048547 -.0073875
s01a_02 2.Mujer	-.0322658	.0074953	-4.30	0.000	-.0469578 -.0175739
logedad_inicio	-.290995	.0303238	-9.60	0.000	-.3504341 -.231556
_cons	3.040256	.1222748	24.86	0.000	2.80058 3.279933

Instrumented: logexp

Instruments: logaest 2.area 2.s01a_02 logedad_inicio

Aquí se reemplaza la experiencia observada por la **experiencia instrumentada** obtenida en la primera etapa. Esto permite estimar el efecto causal de la experiencia sobre los ingresos laborales, corrigiendo la endogeneidad. Los resultados muestran que la experiencia instrumentada tiene un impacto positivo y significativo sobre los ingresos (coeficiente 2.718), la educación también incrementa los ingresos (1.842), mientras que residir en área rural (-0.197) o ser mujer (-0.234) reduce significativamente el ingreso esperado. El instrumento

cumple su función al **explicar la variación de la experiencia sin afectar directamente los ingresos**, permitiendo obtener estimaciones consistentes.

Para la regresión final, se emplea el comando robust, con este comando la regresión sigue siendo la segunda etapa, pero los errores estándar se ajustan para heterocedasticidad; esto **no cambia los coeficientes**, solo hace que los valores de t/z y P sean más confiables. Tal como se muestra a continuación:

```
. ivregress 2sls logy1ab (logexp = logedad_inicio) logaest i.area i.s01a_02, robust
```

Instrumental variables 2SLS regression

Number of obs = 12,906
Wald chi2(4) = 900.18
Prob > chi2 = 0.0000
R-squared =
Root MSE = 1.2131

loglab	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]
logexp	2.717665	.3376243	8.05	0.000	2.055934 3.379397
logaest	1.842055	.1044286	17.64	0.000	1.637379 2.046731
area Rural	-.1967051	.0367539	-5.35	0.000	-.2687415 -.1246687
s01a_02	-.2343348	.0237962	-9.85	0.000	-.2809745 -.187695
2.Mujer	-5.848912	1.387223	-4.22	0.000	-8.567819 -3.130005
_cons					

Instrumented: logexp

Instruments: logaest 2.area 2.s01a_02 logedad_inicio

Variable	Coefficiente	Interpretación
logexp (experiencia instrumentada)	2,718	Un 1% más de experiencia laboral incrementa los ingresos en 2,7% . Representa el efecto causal de la experiencia una vez corregida la endogeneidad.
logaest (años de educación)	1,842	Un 1% más de educación aumenta los ingresos en 1,84% , manteniendo constantes los demás factores.
Área Rural	-0,197	Residir en área rural reduce los ingresos en 19,7% respecto a áreas urbanas, controlando por educación, experiencia y género.
Mujer	-0,234	Ser mujer reduce los ingresos en 23,4% respecto a los hombres, controlando las demás variables.
Constante (_cons)	-5,849	Nivel base de log-ingresos cuando todas las variables independientes son cero (interpretación económica directa limitada).

Respecto a la capacidad predictiva del modelo, podemos afirmar que existe significancia global en el modelo con variables instrumental, debido a que el estadístico F de Fisher tiene un p-value igual a cero, lo que lo haría estadísticamente significativo al 99% de confianza, también podemos afirmar que existe significancia individual debido a que los p-values del estadístico t de student son menores al

0,01 y al 0,05, implicando significancia individual.

En regresiones 2SLS, el R^2 convencional no se reporta ni es confiable, por lo que la interpretación se centra en los coeficientes y su significancia, mostrando la magnitud y dirección de los efectos de cada variable. La edad de inicio laboral se utiliza como instrumento para la

experiencia, porque determina la experiencia acumulada a lo largo de la vida laboral pero no afecta directamente los ingresos actuales más allá de ese efecto, cumpliendo la condición de validez del instrumento y permitiendo estimar un efecto causal consistente de la experiencia sobre los ingresos. El análisis muestra que la experiencia laboral y la educación incrementan los ingresos, mientras que vivir en área rural o ser mujer reduce significativamente los ingresos. Gracias al instrumento, los resultados reflejan relaciones causales más confiables, confirmando que experiencia y educación son los principales determinantes de los ingresos laborales en la muestra. Asimismo, podemos afirmar que las variables en estudio, presentan los signos esperados a priori. Respecto a la capacidad predictiva del modelo, podemos afirmar que existe significancia global en el modelo con variables instrumental, debido a que el estadístico F de Fisher tiene un p-value igual a cero, lo que lo haría estadísticamente significativo al 99% de confianza, también podemos afirmar que existe significancia individual debido a que los p-values del estadístico t de student son menores al 0,01 y al 0,05, implicando significancia individual. Asimismo, podemos afirmar que las variables en estudio, presentan los signos esperados a priori.

• Edad de Inicio Laboral

Un instrumento plausible es la edad de inicio laboral. Teóricamente, bajo el marco de capital humano (Becker-Mincer), comenzar a trabajar más temprano incrementa, ceteris paribus, el tiempo efectivo de experiencia acumulada, por lo que la edad de inicio está fuertemente correlacionada con la experiencia (condición de relevancia). En un modelo de dos etapas, esta variable permite predecir la parte de la experiencia que no está contaminada por factores no observados (motivación, habilidad, shocks idiosincráticos), generando una medida instrumentada de experiencia para la segunda etapa.

Además, la restricción de exclusión es defendible: una vez que controlas por educación, sexo y área de residencia, la edad a la que se comenzó a trabajar no debería afectar directamente los ingresos actuales, sino solo a través de la experiencia acumulada (condición de exogeneidad). En la práctica, la defensa empírica se complementa reportando la fortaleza del instrumento en la primera etapa (F alto) y, si se usan múltiples instrumentos, con tests de sobreidentificación (Hansen/Sargan). Así, la edad de inicio laboral es un instrumento teórica y empíricamente sólido para corregir la endogeneidad de la experiencia.

- **Distancia desde el hogar familiar hasta a institución educativa más cercana**

Otro instrumento para la **experiencia laboral** podría ser la **distancia desde el hogar familiar hasta la institución educativa más cercana**. La lógica es que quienes vivían más lejos de escuelas o universidades podían retrasar su entrada al mercado laboral, afectando así la **experiencia acumulada**, pero esta distancia no debería tener un efecto directo sobre los ingresos actuales, una vez controladas variables como educación, sexo y área de residencia. Esto garantiza la **relevancia**, ya que está correlacionada con la experiencia laboral, pero no con el error en la ecuación de ingresos.

Respecto a la **exogeneidad**, la ubicación geográfica de la vivienda familiar es, en general, independiente de factores individuales que afectan los ingresos futuros. Su impacto se da únicamente a través de la oportunidad o demora en acumular experiencia laboral. Por lo tanto, usar la distancia al centro educativo como instrumento permite corregir el sesgo de endogeneidad de la experiencia de manera consistente y teóricamente defendible.

Referencias Bibliográficas

- INE. (2024). Encuesta de Hogares 2023 – Base de Datos. Instituto Nacional de Estadística de Bolivia. <https://www.ine.gob.bo>
- Organización Internacional del Trabajo (OIT). (2018). La mujer en el mundo del trabajo: Retos y oportunidades hacia la equidad de género. OIT. https://www.ilo.org/global/publications/WCMS_619577/lang-es/index.htm
- Banco Mundial. (2023). Pobreza y equidad en Bolivia: Informe de análisis estructural. Banco Mundial. <https://www.worldbank.org/es/country/bolivia/publication>
- Canelas, C. (2014). The impact of education on labor market outcomes in Bolivia. *International Journal of Educational Development*, 39, 155–165. <https://doi.org/10.1016/j.ijedudev.2014.07.002>
- Lustig, N., & Pessino, C. (2014). Social spending and income redistribution in Latin America: Bolivia, Brazil, Chile, Mexico, Peru and Uruguay. *Public Finance Review*, 42(3), 287–303. <https://doi.org/10.1177/1091142113496134>
- CEPAL. (2022). Panorama Social de América Latina 2022. Comisión Económica para América Latina y el Caribe. <https://www.cepal.org/es/publicaciones/48559-panorama-social-america-latina-2022>

- J. M., F. Kramarz, D. Margolis y K. R. Troske (2001), —The relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States||, *Journal of the Japanese and International Economies* 15: 419 – 436.
- Abowd, J. M., F. Kramarz y D. Margolis (1999), —High Wage Workers and High Wages Firms||, *Econometrica* 67(2): 251 – 333.
- Andersen, L. E. (2001), —Wage Differentials between Bolivian Cities||, *Economía Política II* (1): 48-62.
- Andersen, L. E. y B. Muriel H. (2007), —Informality and Productivity in Bolivia: A Gender Differentiated Empirical Analysis||, *Banco Mundial*, mimeo.
- Aydemir, A. y M. Skuterud (2008), —The Immigrant Wage Differential within and across Establishments||, *Industrial and Labor Relations Review* 61(3): 333-352.
- Bernard, B. A. y J. B. Jensen (1995), “Exporters, Jobs, and Wages in U.S. Manufacturing: 1976 – 1987, *Brookings Papers: Microeconomics*: 67-119.
- Berndt, E.R. y L.R.Christensen (1973), The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929–68, *Journal of Econometrics* 1: 81–114.
- Breau, S. y D. L. Rigby (2006), Is There Really an Export Wage Premium? A Case Study of Los Angeles Using Matched Employee – Employer Data||, *International Regional Science Review* 29(3): 297-310.
- Binswanger, H. P (1974), The Measurement of Technical Change Biases with Many Factors of Production, *American Economic Review* 64(6): 963-976.
- El-Attar, M. y E. Lopez-Bazo (2008), Firm Size and the Entire Distribution of Wages: A Semiparametric Approach, *Applied Economics Letter* 15: 427-430.
- Evia, T. (2010), Eficiencia técnica del sector manufacturero en Bolivia: Una aproximación de frontera de producción estocástica, *Maestrías para el Desarrollo*, Universidad Católica Boliviana —San Pablo, mimeo.