

Construcción de un Pseudopanel con Encuestas de Hogares para el Análisis de las Transiciones de Pobreza en Bolivia *

Gilmar Belzu & Carlos Pantoja **

Resumen:

Este estudio tiene como objetivo medir las transiciones de pobreza en Bolivia y evaluar la viabilidad de los pseudopaneles como herramienta metodológica, dada la falta de datos longitudinales. Para ello, se agrupan individuos en cohortes definidas por características invariantes con el fin de analizar las probabilidades de transición entre los estados de pobreza y no pobreza. Los resultados muestran una movilidad limitada de ingresos, con una persistencia significativa de la pobreza. Además, el nivel educativo y la ubicación geográfica se destacan como factores relevantes en la dinámica de los ingresos. Este trabajo contribuye a la evaluación de los pseudopaneles como una alternativa viable para el análisis de transiciones de pobreza en contextos con datos limitados.

Clasificación JEL: C33, D31 , I32, C81 .

Palabras clave: pseudopaneles, pobreza, encuestas de hogares.

* El contenido del presente documento es de responsabilidad de los autores y no compromete la opinión de Fundación ARU

** Agradecimientos a Wilson Jiménez y a Guillermo Gómez por sus comentarios y recomendaciones.

Comentarios y sugerencias son bienvenidos a: gbelzu@aru.org.bo y cpantoja@aru.org.bo

Abstract:

This study aims to measure poverty transitions in Bolivia and assess the viability of pseudopanel data as a methodological tool, given the lack of longitudinal data. To achieve this, individuals are grouped into cohorts defined by invariant characteristics in order to analyze the probabilities of transition between poverty and non-poverty states. The results show limited income mobility, with significant persistence of poverty. Furthermore, educational level and geographic location emerge as key factors in income dynamics. This work contributes to the evaluation of pseudopanel data as a viable alternative for analyzing poverty transitions in contexts with limited data.

JEL Classification: C33, D31, I32, C81.

Keywords: Pseudopanel data, Poverty, Household surveys

1. Introducción

El análisis de fenómenos socioeconómicos a lo largo del tiempo, como la evolución de la pobreza, la movilidad de ingresos laborales y los efectos de políticas públicas, requiere información longitudinal que permita observar cambios y dinámicas intertemporales. Sin embargo, en muchos países, especialmente en desarrollo, la disponibilidad de paneles de datos es limitada debido al alto costo de recolección, problemas de retención de participantes y falta de continuidad en las encuestas. Este vacío de datos representa un desafío significativo para la investigación y la formulación de políticas basadas en evidencia, ya que las encuestas transversales, aunque ampliamente disponibles, no permiten captar las trayectorias individuales ni explorar relaciones causales en el tiempo.

En este contexto, los pseudopaneles se presentan como una solución metodológica innovadora al transformar datos transversales repetidos en una estructura que simula paneles longitudinales, mediante la agrupación de individuos en cohortes definidas por características invariantes, como la edad o el lugar de nacimiento. A diferencia de los paneles de datos tradicionales, que siguen a las mismas unidades (personas, empresas, hogares, etc.) a lo largo del tiempo, los pseudopaneles agrupan a los individuos en cohortes definidas por características observables y estables, como la edad, el nivel educativo, el sector económico o la región de residencia, las cuales permanecen constantes a lo largo del tiempo. Este enfoque permite realizar análisis dinámicos sin necesidad de contar con datos individuales de panel, lo que resulta especialmente útil cuando existen limitaciones en la disponibilidad de datos o restricciones de privacidad.

El objetivo de este estudio es evaluar las transiciones de pobreza en Bolivia mediante la utilización de un pseudopanel, con el fin de identificar los factores clave que influyen en dichas transiciones y analizar sus dinámicas a lo largo del tiempo. A su vez, se busca explorar el potencial del pseudopanel como herramienta para estudiar los cambios en la pobreza en contextos de datos limitados, como el de Bolivia, y es crucial determinar si realmente puede ser una herramienta efectiva para analizar estos cambios en dichos contextos. En este marco, se examinan los principales factores explicativos que afectan los ingresos de los hogares, tales como la edad, el nivel educativo y la ubicación geográfica. Los resultados obtenidos no solo enriquecen la comprensión de las transiciones de pobreza en Bolivia, sino que también ofrecen insumos clave para el diseño de políticas públicas más inclusivas y efectivas en el contexto socioeconómico del país.

En la siguiente sección se presentará una revisión detallada de la literatura relevante, destacando los avances metodológicos en la construcción de pseudopaneles y su aplicación en contextos similares al de Bolivia. Posteriormente, se describirá la metodología empleada en este estudio, incluyendo el

proceso de armonización de datos y la construcción de cohortes sintéticas. Finalmente, se discutirán los resultados obtenidos, su relevancia para el análisis de la movilidad socioeconómica y las implicancias para el diseño de políticas públicas enfocadas en la reducción de la pobreza y la desigualdad.

2. Revisión de literatura

Los estudios sobre la pobreza y la movilidad social han evolucionado significativamente, pasando de enfoques estáticos a análisis dinámicos que buscan comprender cómo las personas cambian de situación socioeconómica a lo largo del tiempo. Balderrama y Aliaga (2020) en *Feminización de la pobreza en Bolivia*, subraya que la pobreza afecta de manera diferente a distintos grupos, siendo algunos más vulnerables que otros. El análisis desde una perspectiva de género revela que las mujeres a menudo enfrentan mayores dificultades para alcanzar condiciones de vida adecuadas, tanto en términos monetarios como no monetarios. Este documento destaca la importancia de considerar la construcción cultural de las diferencias entre hombres y mujeres al estudiar la pobreza, un factor que los análisis tradicionales de pobreza pueden pasar por alto.

Por otro lado, Ferreira y cols. (2013) abordan la movilidad económica en América Latina, enfocándose en la movilidad intrageneracional e intergeneracional. Este estudio utiliza una combinación de fuentes de datos, desde encuestas de hogares hasta pruebas de evaluación de estudiantes, para entender las transformaciones sociales en la región. Se destaca que la movilidad ascendente intrageneracional ha sido notable, lo que significa que muchas personas han mejorado sus ingresos a lo largo de su vida. Sin embargo, la movilidad intergeneracional sigue siendo baja, indicando que las oportunidades de las personas todavía están influenciadas por el nivel socioeconómico de sus padres.

La metodología de pseudopaneles es una herramienta clave utilizada en ambos estudios para analizar la movilidad a largo plazo. El documento de Balderrama y Aliaga (2020) menciona el uso de pseudo paneles en estudios anteriores para Bolivia, como los de Ferreira y cols. (2013), Vakis et al. (2015) y el PNUD (2016), que muestran avances en la disminución de la pobreza pero también la persistencia de hogares que no pueden salir de esta condición. Estos análisis utilizan datos de encuestas de hogares repetidas en el tiempo para construir una representación dinámica de las trayectorias de los individuos, aunque presentan limitaciones, como la necesidad de imponer supuestos estructurales para generar medidas de movilidad. El documento de Ferreira y cols. (2013) profundiza en la aplicación de esta técnica, y señala que, al no existir datos de panel que sigan a los mismos individuos durante largos períodos, se utilizan los pseudopaneles para calcular la movilidad direccional de los ingresos de manera conservadora, ya que se estiman valores mínimos de movilidad. Esta metodología permite analizar la movilidad de in-

gresos a lo largo del ciclo vital de una generación y desglosarla en “ganadores” y “perdedores”, así como por la clase social inicial de cada hogar.

La metodología de pseudopaneles, como la propuesta por Dang y cols. (2014), utiliza datos de encuestas de hogares de diferentes años para crear un pseudopanel. Esta técnica se basa en la imputación fuera de la muestra, prediciendo el ingreso para los mismos hogares en diferentes períodos, creando así un pseudopanel que simula datos longitudinales. Se utilizan regresiones lineales para modelar el ingreso en función de características observables y constantes en el tiempo, tales como sexo, lugar de nacimiento o educación. Se generan estimaciones de los límites inferior y superior del ingreso, que permiten calcular cotas de movilidad. Este enfoque supera los problemas de atrición¹ que afectan a los datos de panel tradicionales.

La literatura sobre pseudopaneles ha experimentado un crecimiento notable, pero persisten importantes brechas, especialmente en su aplicación en países en desarrollo y en contextos específicos como Bolivia. Aunque los pseudopaneles han demostrado ser útiles en la investigación de fenómenos socioeconómicos en ausencia de datos longitudinales completos, su aplicación en América Latina y economías emergentes aún es limitada. Se requiere más investigación que explore cómo adaptar esta metodología a las particularidades de estas economías, como la alta informalidad laboral y la heterogeneidad de los datos. Es crucial también profundizar en aspectos metodológicos, como el tamaño de las cohortes y los supuestos subyacentes en los análisis, así como en la validación y robustez de los resultados en diferentes contextos.

3. Metodología

La metodología del pseudopanel se emplea como una herramienta clave para analizar las transiciones de pobreza en Bolivia, permitiendo superar las limitaciones inherentes a la falta de datos longitudinales en el país. Este enfoque no solo busca identificar los factores clave que influyen en las dinámicas de pobreza, sino también evaluar la efectividad del pseudopanel como herramienta metodológica en contextos de datos limitados. La metodología propuesta incluye la construcción de cohortes sintéticas y la estimación de probabilidades de transición, que reflejan las dinámicas socioeconómicas a lo largo del tiempo.

¹La atrición en datos de panel se refiere a la pérdida de unidades de observación a lo largo del tiempo, ya sea por no respuesta, abandono del estudio o cambios en la población analizada, lo que puede generar sesgos en el análisis.

3.1. Datos

Para la construcción del panel de datos se emplearon las Encuestas de Hogares realizadas por el Instituto Nacional de Estadística (INE) de Bolivia, abarcando los periodos 2019 y 2023. Estos años fueron seleccionados estratégicamente debido a su relevancia en el contexto de la pandemia de COVID-19, la cual tuvo profundas repercusiones en los ámbitos socioeconómicos. El año 2019 representa el periodo previo a la pandemia, lo que permite establecer una línea base de las condiciones de vida y las dinámicas socioeconómicas existentes. Por su parte, el año 2023 ofrece una perspectiva post-pandemia, en un contexto de recuperación, lo que posibilita analizar los impactos a mediano plazo de esta crisis global en términos de pobreza. Este enfoque comparativo permite identificar cambios estructurales y diseñar intervenciones políticas más adaptadas a las nuevas realidades.

Estas encuestas, recopiladas anualmente, representan una de las fuentes de datos más completas y confiables sobre las condiciones socioeconómicas de los hogares en Bolivia, proporcionando información detallada sobre temas como empleo, educación, ingresos, salud, y condiciones de vivienda. El INE, como la entidad oficial de estadística en Bolivia, garantiza la calidad y rigor metodológico de estos datos, lo que asegura su validez para el análisis longitudinal y comparativo a lo largo de la serie temporal seleccionada.

Las Encuestas de Hogares del INE están diseñadas para ser representativas a nivel nacional, así como a nivel de áreas urbanas y rurales, y por cada uno de los departamentos del país. Esta cobertura garantiza que la muestra capture las principales características y tendencias de la población boliviana en sus diferentes contextos geográficos y socioeconómicos. Además, la representatividad de la muestra permite realizar comparaciones y análisis diferenciales entre zonas urbanas y rurales, así como entre los distintos departamentos, lo que facilita un entendimiento más preciso de las disparidades regionales y los patrones de cambio en el tiempo. Utiliza un diseño de muestreo complejo en dos etapas, con selección de UPMs (Unidad Primaria de Muestreo) por probabilidad proporcional al tamaño de viviendas y selección sistemática de viviendas en la segunda etapa. Esta información es fundamental para el cálculo de los factores de expansión y la ponderación de las observaciones en el panel.

Para aplicar la metodología mencionada, se utilizará específicamente el módulo de persona, ya que este contiene variables clave como el ingreso del hogar y las líneas de pobreza. Estas variables son fundamentales para evaluar la situación socioeconómica de los hogares y su capacidad para enfrentar situaciones adversas, lo que permite realizar un análisis detallado de las transiciones de pobreza.

3.2. Construcción del pseudopanel

El primer paso consistió en procesar y armonizar las variables necesarias para cada cohorte, asegurando su comparabilidad entre las dos bases de datos. Esto implicó ajustar y estandarizar las variables clave, tanto en términos de nomenclatura como de categorías, para garantizar la coherencia entre las definiciones y las estructuras de los conjuntos de datos, como se detalla en el Anexo A. Dado que muchas de las variables están sujetas a cambios a lo largo del tiempo, ya sea por modificaciones en las preguntas de las encuestas o en las categorías de respuesta, fue necesario procesarlas para que permanecieran constantes en todos los años considerados. Este paso permitió mitigar problemas derivados de la heterogeneidad en los datos y asegurar la validez del análisis longitudinal. Además, se normalizaron los nombres de las variables y se creó una nueva variable anual que indicara el año de la oleada en que se encuestó al hogar. Este enfoque sistemático sentó las bases para la construcción del pseudopanel, que permite analizar tendencias y características de las cohortes seleccionadas a lo largo del tiempo.

Para la construcción de cohortes, primero se definieron las cohortes basadas en el año de nacimiento y la edad de los individuos. Se calculó el año de nacimiento restando la edad del año actual y luego se crearon cohortes de edad agrupando a los individuos en rangos de 3 años. El año de nacimiento es una variable fundamental para definir cohortes, ya que representa una característica estable de los individuos a lo largo del tiempo. Agrupar a los individuos por rangos de edad es una práctica común para crear cohortes de tamaño adecuado y reducir la variabilidad. La edad se utiliza para limitar la muestra de estimación a un rango apropiado. A continuación, se generaron etiquetas para cada cohorte de edad.

Se incorporaron también las variables de género, área geográfica (urbano o rural), y departamento, creando nuevas variables para cada una de estas características. El género es otra característica que no varía con el tiempo y que puede influir en la dinámica de la pobreza y otros resultados socioeconómicos. La inclusión de género permite analizar posibles diferencias entre hombres y mujeres dentro de cada cohorte. La ubicación geográfica puede influir en las oportunidades económicas y sociales de los individuos. Incluir el área geográfica permite controlar por estas diferencias y crear cohortes más homogéneas. La inclusión del departamento como variable permite capturar la heterogeneidad regional dentro de un país. Dado que las condiciones económicas y sociales pueden variar significativamente entre departamentos, esta variable ayuda a formar cohortes más específicas y relevantes. Luego, se combinó toda esta información para formar un ID único para cada cohorte, que incluía la cohorte de edad, género, área geográfica y departamento.

3.3. Modelo Econométrico

Las transiciones de pobreza se evaluaron utilizando probabilidades conjuntas y condicionales, estimadas mediante los métodos de límites superiores e inferiores propuestos por Dang y cols. (2014). Este enfoque considera dos escenarios extremos de correlación entre los términos de error: máxima movilidad (correlación nula) y mínima movilidad (correlación perfecta). Estas estimaciones permiten identificar cambios en los estados de pobreza de los hogares, proporcionando una visión más completa de las dinámicas socio-económicas. Por ejemplo, el límite superior para la probabilidad de que un hogar sea pobre en 2019 pero no en 2023 (transición de pobre a no pobre) refleja un escenario donde no existe persistencia en los factores que afectan los ingresos. Por otro lado, el límite inferior supone una completa rigidez estructural, donde las condiciones iniciales determinan completamente las transiciones.

A continuación se presenta el proceso:

Estimación mediante OLS

1. **Modelo por ronda de encuesta:** Se estima una regresión de mínimos cuadrados ordinarios (OLS) en cada ronda de encuesta, con el ingreso como variable dependiente y las características invariantes como variables explicativas:

$$y_{it} = \alpha + \beta X_{it} + \varepsilon_{it},$$

donde:

- y_{it} es el ingreso del hogar i en el año t .
 - X_{it} es un vector de características observables.
 - β son los coeficientes estimados.
 - ε_{it} es el término de error.
2. **Predicción en la segunda ronda:** Los coeficientes estimados en la primera ronda ($\hat{\beta}_1$) se utilizan para predecir los ingresos en la segunda ronda, bajo el supuesto de que las relaciones estructurales entre las características y el ingreso se mantienen constantes:

$$\hat{y}_{i2} = \hat{\alpha}_1 + \hat{\beta}_1 X_{i2}.$$

Límite Superior (Upper Bound)

Se asume que los errores en diferentes períodos son completamente independientes, es decir, la correlación entre ellos es cero: $\text{corr}(e_{i1}, e_{i2}) = 0$.

Procedimiento (No Paramétrico):

1. Se estima la ecuación de ingreso para el período 1, utilizando los datos de la encuesta de ese período. De esta estimación se obtienen los parámetros ($\hat{\beta}_1$) y los residuos ($\hat{\epsilon}_{i1}$).
2. Para cada hogar en el período 2, se toma una muestra aleatoria con reemplazo de los residuos estimados del período 1 ($\hat{\epsilon}_{i1}$), denotándolos como $\tilde{\epsilon}_{i1}$.
3. Se utiliza la ecuación estimada del período 1, junto con los parámetros estimados ($\hat{\beta}_1$), las características del hogar en el período 2 (x_{i2}) y el error simulado $\tilde{\epsilon}_{i1}$ para predecir el ingreso del hogar i en el período 2, proyectado al período 1. Esto se expresa formalmente como:

$$\hat{y}_{i1}^U = \hat{\beta}'_1 x_{i2} + \tilde{\epsilon}_{i1}$$

donde:

- \hat{y}_{i1}^U es el límite superior del ingreso estimado del hogar i del período 2 para el período 1.
 - $\hat{\beta}_1$ son los parámetros estimados de la ecuación de ingresos del período 1.
 - x_{i2} son las características del hogar i en el período 2.
 - $\tilde{\epsilon}_{i1}$ es el error simulado del hogar i tomado de la distribución de errores del período 1.
4. Este procedimiento se repite R veces (generalmente mediante bootstrapping), y se calcula el promedio de las estimaciones obtenidas para obtener el límite superior final.

La probabilidad de que un hogar esté en pobreza en el período 1 y no en el período 2 se calcula como:

$$\hat{P}^{2U}(y_{i1} < z_1 \text{ y } y_{i2} > z_2) = \Phi_2 \left(\frac{z_1 - \hat{\beta}'_1 x_{i2}}{\hat{\sigma}_{\epsilon 1}}, \frac{-z_2 - \hat{\beta}'_2 x_{i2}}{\hat{\sigma}_{\epsilon 2}}, -\rho_S \right)$$

donde Φ_2 es la función de distribución normal bivariada, ρ_S es el valor mínimo de ρ , y $\hat{\sigma}_{\epsilon 1}$ y $\hat{\sigma}_{\epsilon 2}$ son las desviaciones estándar de los residuos de los períodos 1 y 2, respectivamente.

Límite Inferior (Lower Bound)

Se asume que los errores en diferentes períodos están completamente correlacionados, es decir, que la correlación entre ellos es uno: $\text{corr}(e_{i1}, e_{i2}) = 1$.

Procedimiento (No Paramétrico):

1. Se estiman las ecuaciones de ingreso para ambos períodos (1 y 2), obteniendo los parámetros estimados ($\hat{\beta}_1$ y $\hat{\beta}_2$) y los residuos estimados ($\hat{\epsilon}_{i1}$ y $\hat{\epsilon}_{i2}$), así como las desviaciones estándar $\hat{\sigma}_{\epsilon 1}$ y $\hat{\sigma}_{\epsilon 2}$.

2. Se calcula un término de ajuste $\gamma = \sqrt{\frac{\text{Var}(\epsilon_{i1})}{\text{Var}(\epsilon_{i2})}} = \frac{\hat{\sigma}_{\epsilon 1}}{\hat{\sigma}_{\epsilon 2}}$.

3. Se utilizan los parámetros del período 1 ($\hat{\beta}_1$), las características del hogar en el período 2 (x_{i2}) y los residuos estimados del período 2 ($\hat{\epsilon}_{i2}$) para predecir el ingreso del hogar i del período 2 proyectado al período 1. Esto se expresa como:

$$\hat{y}_{i1}^L = \hat{\beta}'_1 x_{i2} + \gamma \hat{\epsilon}_{i2}$$

donde:

- \hat{y}_{i1}^L es el límite inferior del ingreso estimado del hogar i del período 2 para el período 1.
- γ es el término de ajuste, que escala el error del período 2 para hacerlo comparable con el del período 1.
- $\hat{\epsilon}_{i2}$ es el término de error del período 2.

La probabilidad de que un hogar esté en pobreza en ambos períodos se calcula como:

$$\hat{P}^{2L}(y_{i1} < z_1 \text{ y } y_{i2} < z_2) = P(y_{i2} < z_2) - P(y_{i1} > z_1 \text{ y } y_{i2} < z_2)$$

La metodología descrita no solo permite analizar transiciones de pobreza en Bolivia, sino también evaluar la utilidad del pseudopanel como una herramienta metodológica en contextos con datos limitados. En las siguientes secciones se presentan los resultados obtenidos, destacando tanto las dinámicas observadas como las implicaciones de los hallazgos para la investigación y la política pública.

3.4. Supuestos y limitaciones

El método del pseudopanel para estimar las transiciones de pobreza, es decir, los movimientos de individuos o hogares dentro y fuera de la pobreza a lo largo del tiempo, se basa en dos supuestos importantes que determinan los límites superior e inferior de dichas transiciones (Dang y cols., 2014).

Supuesto 1: Estabilidad de la población de referencia

- Se asume que la población de la cual se extraen los datos de las encuestas transversales permanece constante entre las diferentes rondas de la encuesta.
- **Ejemplo:** Si se están estudiando las transiciones de pobreza entre 2010 y 2015, se asume que no hubo migraciones masivas o cambios demográficos significativos que alteren la composición de la población durante ese período.

- **Implicaciones:** Si este supuesto no se cumple, las estimaciones de las transiciones pueden ser inexactas. Por ejemplo, si hubo una gran afluencia de inmigrantes pobres entre 2010 y 2015, la pobreza podría parecer que ha aumentado, aunque los individuos originales de la población no hayan experimentado un empeoramiento de sus condiciones de vida.

Supuesto 2: Correlación positiva de los términos de error

- Se asume que los términos de error en el modelo de ingresos, que representan factores no observados que afectan los ingresos, están positivamente correlacionados entre las rondas de la encuesta.
- **Ejemplo:** Si un hogar experimenta un shock negativo en un año (por ejemplo, la pérdida del empleo del jefe del hogar), es probable que este shock tenga un impacto persistente en los ingresos del hogar en los años siguientes.
- **Implicaciones:** Este supuesto implica que la movilidad de la pobreza, es decir, la facilidad con la que los individuos o hogares pueden salir o entrar en la pobreza, es limitada. Si los errores no estuvieran correlacionados, la movilidad sería mucho mayor, ya que los shocks serían transitorios y no tendrían efectos a largo plazo.

Relación entre los supuestos y los límites

- **Límite superior (sin correlación):** En este caso extremo ($\rho = 0$), se asume que no hay correlación entre los términos de error, lo que implica máxima movilidad. Este límite sobreestima la movilidad real, ya que en la práctica los errores suelen estar correlacionados.
- **Límite inferior (correlación perfecta):** En este otro extremo ($\rho = 1$), se asume que los errores están perfectamente correlacionados, lo que implica mínima movilidad. Este límite subestima la movilidad real.

Las estimaciones de las transiciones de pobreza obtenidas a partir del método del pseudopanel se encuentran entre estos dos límites.

Comprender estos dos supuestos es fundamental para interpretar correctamente los resultados del método del pseudopanel. Es importante tener en cuenta que los límites superior e inferior son solo eso: límites. La movilidad real de la pobreza se encuentra en algún punto entre ellos.

Además de los supuestos, existen limitaciones inherentes que deben ser consideradas:

1. Consistencia de los datos entre años: Los cambios metodológicos en las encuestas del INE pueden haber introducido variaciones en las mediciones, afectando la comparabilidad directa entre 2019 y 2023.

2. Representatividad poblacional: Algunas cohortes, especialmente en áreas rurales o en regiones menos densamente pobladas como Pando y Beni, pueden estar subrepresentadas en el análisis.
3. Suposición de estabilidad poblacional: Este análisis asume que las características clave de las cohortes (edad, género, área) permanecen constantes en el tiempo. Sin embargo, migraciones internas o cambios en las condiciones socioeconómicas podrían haber alterado estas dinámicas.
4. Efectos no observados: Aunque el modelo econométrico intenta controlar factores clave, variables no incluidas, como el acceso a tecnología o infraestructura, pueden influir significativamente en los ingresos.

4. Resultados

Cuadro 1: Modelo de Ingresos Pseudopanel

Variable dependiente: Ingreso por Hogar	2019	2023
Año de Nacimiento	-20.542*** (2.205)	-20.242*** (1.956)
Mujer	-56.256 (41.380)	-60.533 (41.860)
Área	-194.853*** (61.388)	-438.901*** (46.898)
Departamento	47.458*** (9.210)	72.054*** (8.734)
Nivel Educativo 1	388.739*** (69.433)	101.512 (141.552)
Nivel Educativo 2	735.515*** (89.457)	389.906*** (142.024)
Nivel Educativo 3	1,906.540*** (98.255)	1,443.210*** (147.915)
Nivel Educativo 4	1,135.270*** (157.826)	678.133*** (336.855)
Consumo	41,403.456*** (4,313.681)	41,476.412*** (3,858.331)
R^2	0.146	0.102
N	9,160	9,601

Notas: *** p<0.01; ** p<0.05; * p<0.1

El análisis del modelo de ingresos por hogar basado en un pseudopanel resalta varios factores clave que inciden en los ingresos de los hogares en Bolivia. En primer lugar, la edad del jefe de hogar tiene un efecto negativo y significativo, indicando que a medida que aumenta la edad, los ingresos tienden a disminuir.

Aunque la variable de género muestra una tendencia hacia ingresos más bajos para las mujeres, no es estadísticamente significativa en este modelo,

lo que sugiere que otros factores pueden ser más determinantes. En cuanto a la ubicación geográfica, los hogares rurales tienen ingresos considerablemente más bajos que los urbanos, con la disparidad ampliándose en 2023, lo que destaca las desigualdades económicas entre estas zonas.

La ubicación departamental también juega un papel crucial, ya que los hogares en departamentos más desarrollados reportan ingresos más altos, reflejando las disparidades regionales en el país. El nivel educativo del jefe de hogar emerge como uno de los factores más importantes: los hogares con un mayor nivel educativo tienen ingresos significativamente mayores, subrayando la importancia de la educación en la movilidad económica.

Por último, la relación positiva entre consumo e ingresos confirma que los hogares con mayor capacidad de consumo tienen también mayores ingresos, indicando que un ingreso más alto facilita una mayor capacidad para consumir.

A continuación se presentan los resultados de las estimaciones de movilidad de pobreza utilizando el método de pseudopanel, basados en estimaciones de límites no paramétricas (bounds). La tabla muestra tanto las probabilidades conjuntas como las condicionales para el período 2019-2023, desglosadas en sus respectivos límites inferiores y superiores.

Cuadro 2: Estimaciones de límites 2019 - 2023

Probabilidades conjuntas no paramétricas	Límite Inferior	Límite Superior
Pobre, pobre	0.277	0.126
Pobre, no pobre	0.035	0.174
No pobre, pobre	0.025	0.176
No pobre, no pobre	0.663	0.524
N	10,062	10,062

Estimado usando una línea de pobreza de 838.2789 en el año 1 y 865.0668 en el año 2.

Las filas muestran la fracción de la población en el rango de edad seleccionado que está en cada una de las cuatro categorías. Por ejemplo, 'Pobre, pobre' indica la fracción que fue pobre en el año 1 y pobre en el año 2.

Las probabilidades presentadas reflejan la fracción de la población que experimenta transiciones entre diferentes combinaciones de estados de pobreza y no pobreza. En cuanto a la interpretación de las filas, para el estado "Pobre, pobre", la probabilidad de que una persona que era pobre en 2019 siga siendo pobre en 2023 varía entre un límite inferior de 0.277 (27.7 %) y un límite superior de 0.126 (12.6 %).

En el caso de "Pobre, no pobre", la probabilidad de que una persona que era pobre en 2019 deje de ser pobre en 2023 oscila entre 0.035 (3.5 %) y 0.174 (17.4 %). Para "No pobre, pobre", la probabilidad de que una persona que no era pobre en 2019 caiga en la pobreza en 2023 está entre 0.025 (2.5 %) y 0.176 (17.6 %).

Finalmente, para "No pobre, no pobre", la probabilidad de que una persona que no era pobre en 2019 permanezca fuera de la pobreza en 2023 se encuentra entre 0.663 (66.3 %) y 0.524 (52.4 %). Es importante señalar que las probabilidades conjuntas no están condicionadas a ningún estado inicial, ya que representan la distribución global de las transiciones.

Cuadro 3: Probabilidades condicionales de pobreza 2019-2023

Estado en el año 2 dado el estado en el año 1	Límite Inferior	Límite Superior
Pobre Pobre en el año 1	0.888	0.421
No pobre Pobre en el año 1	0.112	0.579
Pobre No pobre en el año 1	0.036	0.251
No pobre No pobre en el año 1	0.964	0.749
N	10,062	10,062

Estimado usando una línea de pobreza de 838.2789 en el año 1 y 865.0668 en el año 2.

Las filas muestran la probabilidad de cada uno de los cuatro estados en el año 2, dado el estado en el año 1.

Estas probabilidades están condicionadas al estado inicial de pobreza o no pobreza en 2019, y muestran la probabilidad de estar en un estado final en 2023. Para la fila "Pobre a pobre", la probabilidad de seguir siendo pobre en 2023, dado que se era pobre en 2019, varía entre un límite inferior de 0.888 (88.8 %) y un límite superior de 0.421 (42.1 %).

En cuanto a la fila "Pobre a no pobre", la probabilidad de salir de la pobreza en 2023, dado que se era pobre en 2019, oscila entre 0.112 (11.2 %) y 0.579 (57.9 %). Para "No pobre a pobre", la probabilidad de caer en la pobreza en 2023, dado que no se era pobre en 2019, está entre 0.036 (3.6 %) y 0.251 (25.1 %).

Por último, para "No pobre a no pobre", la probabilidad de permanecer fuera de la pobreza en 2023, dado que no se era pobre en 2019, se encuentra entre 0.964 (96.4 %) y 0.749 (74.9 %).

Es importante destacar las diferencias entre las probabilidades conjuntas y condicionales. Las probabilidades conjuntas representan fracciones de la población total en cada combinación de estados, sin hacer referencia al estado inicial. En cambio, las probabilidades condicionales reflejan probabilidades relativas para individuos dentro de un grupo específico, ya sea aquellos que eran pobres o no pobres en 2019.

El código utilizado para realizar los análisis y generar los resultados se encuentra disponible en un repositorio de GitHub. Para acceder al código completo y a los detalles sobre su ejecución, se puede consultar el Anexo B o escanear el código QR proporcionado en dicho anexo.

5. Discusión y Conclusiones

Esta investigación analiza la relevancia de los pseudopaneles en el estudio de las dinámicas socioeconómicas en contextos donde los datos longitudinales son limitados, como ocurre en Bolivia. Aunque los pseudopaneles ofrecen una alternativa innovadora frente a los estudios de panel tradicionales, sus limitaciones metodológicas deben ser discutidas para comprender su alcance real en la medición de las transiciones de pobreza.

El análisis comparativo con estudios previos, como los de Stampini y cols. (2015) y Ferreira y cols. (2013), evidencia un avance metodológico en la aplicación de pseudopaneles para capturar cambios intertemporales. A diferencia de dichos estudios, que se centraron en la movilidad social o la pobreza, este trabajo evalúa específicamente la viabilidad del uso de un pseudopanel construido con las encuestas de hogares de Bolivia, explorando su potencial como herramienta para el análisis de las transiciones de pobreza.

Una limitación importante identificada es la falta de significancia estadística en algunas cohortes, especialmente en áreas rurales. Esto podría deberse a sesgos en la recolección de datos o a la menor representatividad de estas regiones en la muestra. Además, factores externos como la volatilidad económica y la pandemia de COVID-19 pudieron haber introducido fluctuaciones que no fueron completamente capturadas por el modelo. A pesar de estas limitaciones, los resultados destacan la utilidad de los pseudopaneles en contextos con datos restringidos. A través del análisis de las encuestas de hogares de 2019 y 2023, se identificaron patrones de pobreza y movilidad socioeconómica en un período crítico. Se confirma que el nivel educativo sigue siendo un determinante clave del ingreso, mientras que las disparidades urbano-rurales y de género persisten como desafíos estructurales.

Aunque los pseudopaneles pueden no capturar con total precisión las dinámicas a largo plazo, la metodología utilizada ha demostrado ser una herramienta valiosa para transformar datos transversales en aproximaciones longitudinales. Este enfoque no solo permite identificar cohortes vulnerables que requieren atención específica, sino que también proporciona información clave para el diseño de políticas públicas más ajustadas a las realidades locales. Futuros estudios deberían considerar el impacto de factores adicionales, como la migración y los efectos de programas gubernamentales, en la movilidad socioeconómica. Asimismo, extender el análisis a períodos más largos permitiría una comprensión más profunda de las tendencias intertemporales, fortaleciendo la base empírica para la formulación de políticas basadas en evidencia.

Referencias

- Balderrama, C. C., y Aliaga, G. G. (2020). Feminización de la pobreza en Bolivia: Un análisis a partir de paneles sintéticos. *Revista de Análisis del BCB*, 32(1), 89-122. Descargado de <https://ideas.repec.org/a/blv/journal/v32y2020i1p89-122.html>
- Borysov, S. S., y Rich, J. (2021). Introducing synthetic pseudo panels: application to transport behaviour dynamics. *Transportation*, 48(5), 2493–2520. Descargado 2024-11-06, de <https://doi.org/10.1007/s11116-020-10137-5>
- Bourguignon, F., Goh, C.-c., y Kim, D. I. (2004). *Estimating individual vulnerability to poverty with pseudo-panel data* (Policy Research Working Paper Series n.º 3375). The World Bank. Descargado de <https://ideas.repec.org/p/wbk/wbrwps/3375.html>
- Dang, H.-A., Lanjouw, P., Luoto, J., y McKenzie, D. (2014). Using repeated cross-sections to explore movements into and out of poverty. *Journal of Development Economics*, 107, 112–128. Descargado 2024-11-21, de <https://www.sciencedirect.com/science/article/pii/S0304387813001521>
- Ferreira, F. H. G., Julian, M., Jamele, R., Luis-Felipe, L.-C., María, L., y Renos, V. (2013). *La movilidad económica y el crecimiento de la clase media en américa latina*. Banco Mundial. Descargado de <https://hdl.handle.net/20.500.12799/1406>
- Garcés-Urzainqui, D., Lanjouw, P., y Rongen, G. (2021). Constructing synthetic panels for the purpose of studying poverty dynamics: A primer. *Review of Development Economics*, 25(4), 1803–1815. Descargado 2024-11-06, de <https://onlinelibrary.wiley.com/doi/abs/10.1111/rode.12832>
- Holger, A., Agnes, D., Dana, M., y Anja, S. (2009). Development of a panel dataset from the IAB establishment panel cross-sectional data. *Research Papers in Economics*. Descargado 2024-11-12, de <https://typeset.io/papers/development-of-a-panel-dataset-from-the-iab-establishment-53qka77xhu>
- Mercado, A. F., y Leiton-Quiroga, J. G. M. (2009). The dynamics of poverty in bolivia. *Revista Latinoamericana de Desarrollo Económico*, 7(11), 45–82. Descargado 2024-11-21, de <https://lajed.ucb.edu.bo/a/article/view/a73>
- Stampini, M., Robles, M., Sáenz, M., Ibararán, P., y Medellín, N. (2015). Poverty, vulnerability and the middle class in latin america. *Inter-American Development Bank*. Descargado 2024-11-21, de <https://publications.iadb.org/en/poverty-vulnerability-and-middle-class-latin-america>

Universiteit Amsterdam, y Rongen, G. (2021). Manual for the estimation of a synthetic panel and vulnerability analysis. *Data and Evidence to End Extreme Poverty*. Descargado 2025-01-03, de <https://povertyevidence.org/resources/manual-for-the-estimation-of-a-synthetic-panel-and-vulnerability-analysis/>

Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. En L. Mátyás y P. Sevestre (Eds.), *The econometrics of panel data: Fundamentals and recent developments in theory and practice* (pp. 369–383). Springer. Descargado 2024-11-12, de https://doi.org/10.1007/978-3-540-75892-1_11

A. Documentación técnica adicional

Para la gestión y procesamiento de los datos utilizados en esta investigación, se empleó el programa **iecodebook**. Este software facilita la creación y gestión de diccionarios de variables, así como la documentación y análisis de bases de datos complejas, permitiendo una mejor organización y presentación de los datos. Más información en <https://dimewiki.worldbank.org/Iecodebook>.

B. Código en Github

El código utilizado en esta investigación está disponible en GitHub. Puedes acceder a él escaneando el código QR o mediante el siguiente enlace:



<https://github.com/CarlosPantojaDuran/Pseudopanel-con-las-encuestas-de-hogares-de-Bolivia-para-medir-transiciones-de-pobreza>