

Determinantes del empleo informal en Bolivia: Un análisis conjunto de técnicas econométricas tradicionales y métodos de machine learning

Determinants of informal employment in Bolivia: a combined analysis of traditional econometric techniques and machine learning methods

Ibhar C. Beramendi Illanes*

Universidad Privada Boliviana (UPB), Bolivia

Email: ibharberamendi@upb.edu

 0009-0004-1298-6273

Ivette Illanes Fajardo

Escuela Militar de Ingeniería, Bolivia

ivetteillanes77@gmail.com

 0009-0008-2597-2013

Resumen: Este estudio analiza los determinantes del empleo informal en Bolivia mediante una combinación de técnicas econométricas tradicionales, métodos de machine learning y enfoques híbridos. Utilizando datos de las Encuestas de Hogares 2022 y 2023, se identifican los factores individuales y del hogar que influyen en la probabilidad de pertenecer al empleo informal. Los resultados muestran que variables como la edad, el nivel educativo, el ingreso del hogar y el género son determinantes clave. El Random Forest destaca el papel central de los ingresos laborales, usualmente excluidos por problemas de endogeneidad. El Adaptive Lasso permite identificar relaciones no lineales e interacciones complejas, como las asociadas al género, la pertenencia a grupos originarios y la presencia de niños pequeños en el hogar. Se concluye que el fenómeno del empleo informal responde a dinámicas multidimensionales que requieren enfoques analíticos integradores para el diseño de políticas públicas más efectivas y focalizadas.

Palabras clave: Empleo Informal, Probit, Machine Learning, Adaptive Lasso

Clasificación JEL: J46, C20, C45

Abstract: This study examines the determinants of informal employment in Bolivia by combining traditional econometric techniques, machine learning methods, and hybrid approaches. Using data from the 2022 and 2023 Household Surveys, we identify individual and household-level factors influencing the likelihood of being in informal employment. The results show that variables such as age, education level, household income, and gender are key determinants. Random Forest highlights the central role of labor income, often excluded due to endogeneity concerns. Adaptive Lasso helps identify nonlinear relationships and complex interactions, such as those associated with gender, indigenous group membership, and the presence of young children in the household. We conclude that informal employment is a multidimensional phenomenon requiring integrated analytical approaches for the design of more effective and targeted public policies.

Keywords: Informal Employment, Probit, Machine Learning, Adaptive Lasso

Agradecimientos:

Agradecemos Los Valiosos Comentarios, Sugerencias Y Apoyo Que Contribuyeron Significativamente Al Desarrollo De Este Trabajo, En Especial A Laura Guzmán Navarro, Juan Pablo Barrancos Rojas, Natalia Puch Melgar Y Andrés Viruez Velasco. El Contenido Del Presente Documento Es De Responsabilidad Del (De Los) Autor(Es) Y No Compromete La Opinión De La(S) Entidad(Es) A La(S) Que Pertenece(N) El(Los) Mismo(S).

*Corresponding autor

1. Introducción

En Bolivia, el empleo informal representa un componente estructural de su economía asociado a peores estándares de vida [1], [2]. Identificar sus determinantes, es crucial para diseñar políticas públicas efectivas que promuevan la formalización y mejoren las condiciones laborales Banco Mundial [3].

En la región latinoamericana los principales determinantes del empleo informal se han agrupado en factores individuales y del hogar. La búsqueda de estos determinantes se ha realizado a través de técnicas tradicionales de econometría (TT), como ser modelos logit y probit. Sin embargo, estas técnicas enfrentan limitaciones al capturar interacciones complejas entre variables o patrones no lineales en datos con alta dimensionalidad [4]. En este escenario, se han desarrollado técnicas híbridas apoyadas en el machine learning (ML), rescatando el aporte de esta técnica para predecir comportamientos junto a el enfoque causal que proporcionan las TT.

Si bien las técnicas tradicionales (TT) de econometría y el machine learning (ML) difieren en sus objetivos y enfoques metodológicos, el crecimiento del big data y la necesidad de resolver problemas económicos complejos han motivado la combinación de ambos enfoques metodológicos [5], [6], [7], [8]. No obstante, esta convergencia exige un análisis cuidadoso de sus diferencias, seleccionando su aplicación individual o combinada según el objetivo, para mejorar el análisis empírico mediante sus ventajas comparativas [5], [9], [10]. Estas diferencias metodológicas adquieren especial relevancia en el estudio de fenómenos socioeconómicos multifactoriales, como el empleo informal, un fenómeno ampliamente reconocido como obstáculo clave para el desarrollo en economías emergentes [11].

Este estudio contribuye al análisis de los determinantes del empleo informal en Bolivia mediante TT econométricas, métodos de ML y estrategias híbridas que combinan ambos paradigmas. El objetivo es contrastar los hallazgos según cada enfoque metodológico y generar evidencia robusta para el diseño de políticas públicas más efectivas, así como para mejorar el entendimiento del empleo informal en Bolivia. A través de este enfoque híbrido, se aporta a la literatura de medición de determinantes del empleo informal con una comparación de la efectividad de cada técnica, y aportando a afinar los efectos y causalidad de cada variable sobre la elección de empleo informal.

La utilización de métodos híbridos ha permitido analizar una amplia cantidad de interacciones entre variables y seleccionar aquellas más relevantes, lo cual ha resaltado relaciones de causalidad más complejas, directas e indirectas, que las TT no logran capturar. Entre estas relaciones se encuentra que la edad, años

de estudio e ingresos del hogar tienen una relación no lineal con la probabilidad encontrarse en el sector informal, que el efecto del género (ser mujer) afecta a la probabilidad de ser informal no solo en sí misma sino también a través de la educación (años de estudio), y la pertenencia a un grupo originario; y que, a su vez, pertenecer a un grupo originario no tiene un efecto en sí mismo en la probabilidad de informalidad, pero sí a través del nivel de ingresos del hogar.

Por otro lado, los modelos basados en técnicas de ML muestran una relación importante entre los ingresos laborales y la decisión de informalidad, insinuando una relación más compleja respecto a las expectativas entre estas dos variables más allá de la causalidad inversa. Ahondar en esta relación, podría ayudar a comprender mejor el fenómeno de informalidad en el país o incluso en la región.

El documento se organiza de la siguiente manera: la Sección 2 revisa la literatura relevante, la Sección 3 detalla los datos y metodología, la Sección 4 presenta los resultados y la Sección 5 discute las conclusiones y recomendaciones.

2. Revisión de la literatura

La informalidad es un concepto amplio y complejo que abarca una diversidad de actividades económicas no reguladas por los marcos normativos y regulatorios del Estado. Su definición varía según el enfoque adoptado, desde una perspectiva económica, se asocia con comportamientos de incumplimiento normativo, como la evasión tributaria o la elusión de regulaciones laborales, mientras que desde una visión estadística se vincula a la dificultad de observación y registro en las cuentas nacionales [12]. Aunque la heterogeneidad de las investigaciones complique la definición, una definición ampliamente utilizada por su comparabilidad internacional, es la propuesta por la Organización Internacional del Trabajo, donde la informalidad se define como el conjunto de actividades económicas realizadas por personas o unidades productivas que, en la práctica o en la ley, no están cubiertas por arreglos formales como la legislación laboral, tributaria o de seguridad social [13].

Con dicha definición de empleo informal, es relevante conocer las teorías que explican las causas de porque los agentes económicos optan por ser parte del sector informal, ya que en función a estas se pueden obtener los fundamentos teóricos para comprender el empleo informal e identificar sus determinantes. La teoría del mercado laboral dual conceptualiza al trabajo informal como un sector laboral de segunda donde se ubican los trabajadores excluidos de los mercados primarios (trabajos regulados y que pagan impuestos), caracterizado por la ausencia de regulación y protección laboral, los cuales son un reflejo de

la estratificación racial, por clases y de género en nuestra sociedad [14].

La teoría neoclásica resalta que el pertenecer al sector informal es el resultado de la evaluación racional de los costos y beneficios relativos de entrar al sistema legal; sin embargo, ante fallas o intervenciones de mercado, la informalidad puede actuar como una escapatoria al desempleo [12]. Por otro lado, [15] explica que los empleos informales pueden actuar como una red de seguridad para las familias de bajos ingresos y marginales, ya que facilitan la posibilidad de generar ingresos a aquellos que no tienen acceso al mercado formal. Asimismo, las teorías macroeconómicas explican que la informalidad podría ser el resultado del sistema impositivo, composición del mercado laboral y la debilidad institucional [16].

Igual que la informalidad, el empleo informal es un tema cuya definición presenta desafíos significativos debido a su naturaleza multidimensional y heterogénea, lo que genera diversas definiciones operativas que varían según la dimensión y perspectiva considerada, ya sea contractual, afiliación a la seguridad social, tamaño de la empresa o registro administrativo [17], [18], [19]. Esta diversidad dificulta la comparabilidad entre estudios, regiones y periodos. Por lo tanto, resulta esencial considerar definiciones estándar junto a enfoques sensibles al contexto específico para lograr análisis más robustos y políticas efectivas [17]. En este sentido, la [20] propone una definición teórica ampliamente aceptada:

“El empleo informal puede definirse como cualquier actividad de las personas para producir bienes o prestar servicios a cambio de una remuneración o un beneficio que no esté efectivamente cubierta por disposiciones formales como las leyes comerciales, los procedimientos para declarar las actividades económicas, el impuesto sobre la renta, la legislación laboral y las leyes y reglamentos de seguridad social que proporcionan protección contra los riesgos económicos y personales asociados al ejercicio de las actividades.”

Esta definición servirá como base para definir los criterios operativos aplicados en esta investigación para diferenciar a los trabajadores formales e informales. Por tanto, la definición operativa de empleo informal utilizada en esta investigación es categorizar a los individuos como empleados informales si caen en alguna de las siguientes categorías:

a) Empleados y empleadores con remuneración trabajando en empresas privadas, con 5 o menos personas y que no reciben o no esperan recibir beneficios sociales (subsídios, bonos de natalidad, seguros de salud).

b) Trabajadores por cuenta propia, empleadores sin remuneración y cooperativistas de producción trabajando en instituciones con 5 o menos personas.

c) Trabajadores familiares o aprendices sin remuneración.

d) Empleados del hogar.

Esta definición de empleo informal es similar a medidas tradicionales que se utilizaron para medir la informalidad en Bolivia. De manera general en el caso boliviano la informalidad suele estimarse en base a la perspectiva productiva, donde se categoriza a un empleado informal en base al sector de la economía (público o privado), la clasificación por tipo de trabajo (asalariado, trabajador por cuenta propia, empleador, etc.) y tamaño de la firma (cantidad de trabajadores). Estas mediciones, responden a las recomendaciones de la OIT y suelen presentar estimaciones del tamaño del empleo informal muy similares en el tiempo [21], [22], [23], [24], [25], [26], [27], [28], [29]. No obstante, también se debe mencionar que en algunas ocasiones la informalidad en el país se estima en base a la perspectiva legal, donde se suele considerar el cumplimiento de la normativa laboral como principal criterio y su medición se basa principalmente en información sobre la afiliación al fondo de pensiones, seguros contra accidentes laborales o vacaciones pagadas [1], [21], [30]; sin embargo, a pesar de que no sea la corriente dominante usualmente por la falta de datos y comparabilidad, ambas perspectivas presentan estimaciones similares del empleo informal.

Enfocándonos en el contexto latinoamericano, existen diversas investigaciones que buscaron las determinantes del empleo informal, en los que podemos encontrar patrones muy claros que van acompañados de modelos y justificaciones teóricas que explican la elección de dichas determinantes.

Las principales determinantes del empleo informal con un especial foco en la región suelen identificarse mediante el uso de TT econométricas, específicamente, mediante el uso de modelos de clasificación binaria no lineales (logit y probit). Las determinantes pueden agruparse en factores individuales y del hogar. Entre los factores individuales se encuentran el nivel educativo, el género, la edad, el estado civil, la experiencia laboral, haber migrado recientemente, la pertenencia a determinados grupos sociodemográficos, la existencia de ingresos no laborales y la presencia de alguna discapacidad. En cuanto a los factores del hogar, influyen el nivel de ingresos, los indicadores de riqueza, la presencia de menores y adultos mayores en el hogar, la presencia de trabajadores en el hogar, la residencia en zonas rurales y la situación de pobreza [2], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41].

Para el caso boliviano, las determinantes del empleo informal

que suelen encontrarse son factores asociados a la educación, ya que esta reduce de manera significativa la probabilidad de inserción en el sector informal, mientras que a la vez los retornos de la educación son sustancialmente menores en dicho sector, reforzando la segmentación del mercado de trabajo [42], [43], [44]. Asimismo, se observa una mayor incidencia de informalidad entre las mujeres, vinculada a segregación ocupacional y la concentración femenina en actividades de comercio y servicios de baja productividad [42], [45], [46]. Desde el lado institucional, varios trabajos señalan que la rigidez de la normativa laboral incluyendo el salario mínimo y los costos de despido genera incentivos limitados para la formalización en micro y pequeñas unidades productivas [21], [47], [48].

De esta manera, siguiendo un enfoque parsimonioso y considerando tanto las teorías que explican la decisión de participar en el empleo informal como investigaciones previas que identifican sus determinantes, se analizan como factores relevantes para explicar el empleo informal para Bolivia: la edad, el género, los años de estudio, los ingresos per cápita del hogar, los ingresos no laborales, la cantidad de miembros menores de 5 años en el hogar y la pertenencia a una nación originaria. Asimismo, se incorporan como variables de control el área de residencia (rural o urbana) y el departamento.

Enfocándonos en las diferencias metodológicas, como se presentó anteriormente las TT econométricas han sido ampliamente utilizadas para la identificación de las determinantes del empleo informal, de manera que las investigaciones previas que buscaron encontrar las determinantes del empleo informal se han basado principalmente en modelos de probabilidad no lineal (logit/probit). Brindando la posibilidad de estimar efectos marginales promedio y relaciones causales bajo supuestos paramétricos [49]. Sin embargo, estas técnicas enfrentan limitaciones al capturar interacciones complejas entre variables o patrones no lineales en datos con alta dimensionalidad [4].

Por otro lado, el uso de ML, por su naturaleza, no puede ser utilizado directamente para la identificación de las determinantes o ser interpretado con el enfoque causal con el que se suelen analizar las determinantes [6]. Esto se debe a que los algoritmos de ML están optimizados para maximizar la precisión predictiva y no para estimar parámetros estructurales o efectos causales [50]. No obstante, a pesar de tener importantes limitaciones al momento de identificar determinantes, el ML ha mostrado ser útil al momento de formular políticas públicas al proveer una manera de identificar o predecir comportamientos no deseables por la población o actores económicos que pueden transformarse en oportunidades de política para combatir dichos comportamientos y la optimización de recursos [51], [52].

Como resultado de los beneficios de ambos enfoques se fueron abriendo distintas alternativas que intentan reconciliar la capacidad predictiva de los modelos de ML y el enfoque causal de las TT econométricas [6]. Estas técnicas incluyen, pero no se limitan a Causal Forests, Double/Debiased ML y Adaptive Lasso [53], [54], [55]. En consecuencia, este estudio implementa un enfoque triple para el análisis de las Encuestas de Hogares 2022-2023: un modelo Probit tradicional que preserva la interpretabilidad causal, un algoritmo Random Forest que maximiza la capacidad predictiva capturando no linealidades, y un modelo Adaptive Lasso que permite la selección automática de variables a través de la penalización de la suma de los valores absolutos de los parámetros estimados. Esta estrategia metodológica múltiple permite contrastar los resultados desde distintas perspectivas analíticas, manteniendo el rigor en la identificación de determinantes clave del empleo informal y su potencial traslado a políticas públicas efectivas.

3. Datos y metodología

La fuente de datos utilizada para este estudio corresponde a las Encuestas de Hogares 2022 y 2023, realizadas por el Instituto Nacional de Estadística (INE) durante el segundo semestre de cada año [56], [57]. Estas encuestas tienen como objetivo central la recopilación de datos sobre las principales características demográficas y socioeconómicas de la población boliviana, constituyéndose en una herramienta clave para el análisis del mercado laboral y la estructura de ingresos de los hogares [57]. En particular, la encuesta de hogares proporciona información detallada sobre la condición de actividad, ocupación principal y secundaria, así como los ingresos percibidos tanto por trabajadores asalariados como independientes, aspectos fundamentales para el estudio de la informalidad laboral.

Ambas encuestas utilizan un diseño muestral probabilístico, estratificado y multietápico [56], [57]. El tamaño de muestra planificado fue de 12.816 viviendas en 2022 [56] y de 12.948 en 2023 [57], cerciorando representatividad a nivel nacional, urbano, rural y departamental. La riqueza de estas bases de datos permite un análisis profundo de los determinantes de la informalidad en Bolivia, considerando múltiples dimensiones relacionadas con el empleo y las fuentes de ingreso de la población.

Algunos hechos estilizados del empleo informal en el país se pueden observar en la Tabla 1. El empleo formal se asocia con mejores resultados socioeconómicos, evidenciados por mayores ingresos laborales, mayores ingresos del hogar per cápita, mayor nivel educativo y una menor incidencia de pobreza. Además, quienes se desempeñan en este sector tienden a ser en promedio más jóvenes y trabajar una cantidad de horas similar a quienes están en el empleo informal. Este último, en contraste, no solo

muestra desventajas en términos de ingresos, educación y pobreza, sino que también concentra a poblaciones con menor acceso a oportunidades laborales, lo que podría estar reforzando

dinámicas de exclusión asociadas a características individuales de los empleados informales.

Tabla 1: Hechos Estilizados, Informalidad En Bolivia, 2022 Y 2023

Indicador	EH 2022			EH 2023		
	Formal	Informal	Diferencia	Formal	Informal	Diferencia
Ingresos (BOB/mes)						
Laborales	4558.52	2545.93	2012.59*** (41.51)	3950.95	2544.73	1406.22*** (64.64)
Hogar per cápita	2532.82	1404.52	1128.30*** (25.85)	2322.37	1447.89	874.48*** (27.12)
Horas trabajadas						
Semanales (ocup. principal)	43.37	43.24	0.13 (0.33)	44.88	43.64	1.24*** (0.30)
Características sociodemográficas						
Edad (años)	38.96	40.27	-1.31*** (0.27)	36.86	42.98	-6.12*** (0.23)
Años de estudio	14.04	9.27	4.77*** (0.08)	13.31	8.9	4.41*** (0.07)
Mujeres (%)	39.42	44.43	-5.01*** (0.86)	38.08	46.77	-8.69*** (0.75)
Pertenencia a nación originaria (%)	20.68	33.86	-13.18*** (0.80)	22.01	37.75	-15.74*** (0.70)
Indicadores sociales						
Pobreza (%)	12.57	36.02	-23.45*** (0.79)	14.69	39.58	-24.89*** (0.68)
Trabajo infantil (%)	1.46	8.38	-6.92*** (0.44)	2.69	7.35	-4.66*** (0.35)
Trabajo de ancianos (%)	12.62	20.81	-8.19*** (0.70)	9.96	26.9	-16.94*** (0.61)

***p≤0.01; **p≤0.05; *p≤0.10

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

Como se mencionó en la revisión de la literatura, una TT econométrica clásicamente utilizada para encontrar los determinantes del empleo informal es el modelo probit, el cual servirá como punto de partida y comparación. Este es un modelo es-

tadístico para respuestas binarias en el que la probabilidad de respuesta se obtiene evaluando la función de distribución normal acumulada estándar en una función lineal de las variables explicativas [49].

Su descripción en un contexto econométrico tradicional es:

$$1) P(y = 1|x) = P[e > -(x\beta)|x] = G(x\beta)$$

Donde $P(y = 1 | x)$ es la probabilidad de que, y sea igual a 1, dado los valores de x . e es un término de error no observable. $x\beta$ es una combinación lineal de las regresoras y sus coeficientes. G es la función de distribución normal acumulada estándar y los coeficientes, β se estiman mediante máxima verosimilitud.

En el contexto de esta investigación, se estiman modelos probit para las bases EH 2022 y EH 2023, donde la variable dependiente y representa la condición de pertenencia al empleo informal. Con el objetivo de favorecer la comparabilidad con estimaciones futuras, en particular con el modelo Adaptive Lasso, se opta por normalizar las variables continuas consideradas como posibles determinantes del empleo informal. La variable dependiente toma el valor de $y=1$ para empleo informal y $y=0$ para empleo formal. La matriz de regresores x incluye valores normalizados de edad, años de estudio, ingreso per cápita del hogar, ingreso no laboral y número de personas menores de cinco años. Asimismo, se incorporan variables indicadoras (dummies) para género (referencia: hombres), pertenencia a una nación originaria (referencia: no pertenece), área de residencia (referencia: urbana) y departamento (referencia: La Paz).

En la búsqueda de las determinantes del empleo informal, el objetivo principal es estimar los efectos de x_j sobre $P(y = 1 | x)$; no obstante, los coeficientes presentan únicamente el signo del efecto parcial, para analizar el efecto de las regresoras sobre la probabilidad de ser un empleado informal se necesita información específica de para cada individuo, de manera que se opta por analizar los efectos marginales promedio para capturar el efecto de las regresoras.

Enfocando la discusión en las técnicas de ML, un modelo reconocido por su adaptabilidad y que no requiere un entrenamiento demasiado complejo es el modelo Random Forest. Este es un clasificador compuesto por un conjunto de clasificadores con estructura de árbol, $\{h(x, \theta_k), k = 1, \dots\}$ donde $\{\theta_k\}$ son vectores aleatorios independientes e idénticamente distribuidos. Cada árbol emite un voto unitario por la clase más popular correspondiente en base a x [58]. Su estimación consiste en construir múltiples árboles de decisión utilizando la técnica de *bagging*. Además, en cada nodo de cada árbol, se selecciona aleatoriamente un subconjunto de variables predictoras para determinar la mejor partición mediante un top-greedy algorithm, que evalúa splits basados en el índice Gini. Los hiperparámetros (número de árboles, predictores por nodo, tamaño mínimo de los nodos finales) son ajustados mediante validación cruzada y la predicción final se obtiene mediante votación

mayoritaria entre los árboles del conjunto.

En el contexto de esta investigación, para el modelo Random Forest, la variable dependiente y también representa la condición de pertenencia al empleo informal; sin embargo, las variables independientes utilizadas para entrenar al modelo son diferentes en comparación al modelo probit. Se opta por utilizar un mayor número de variables con el fin de mejorar el entrenamiento y capacidad predictiva del modelo, ya que a diferencia de las TT econométricas no se busca entender las causas o efectos y se pueden obviar aspectos relacionados a la endogeneidad. Las variables son los valores normalizados de la edad, los años de estudio, el ingreso per cápita del hogar, el ingreso laboral, el ingreso no laboral, la brecha de pobreza, la brecha de pobreza extrema, el número total de personas en el hogar, la cantidad de personas menores de 5 años, 18 años y mayores a 54 años, número de personas desempleadas en el hogar y número de personas parte de la población económicamente activa en el hogar. A su vez, se incorporan variables categóricas que capturan dimensiones clave como el género, el estado civil, la pertenencia a una nación originaria, la presencia de alguna discapacidad (solo disponible para la EH 2023), haber migrado en los últimos cinco años (solo disponible para la EH 2023), contar con una ocupación secundaria, la situación de pobreza y pobreza extrema, el área de residencia, el grupo ocupacional principal, y el departamento de residencia.

Los hiperparámetros seleccionados para el modelo Random Forest incluyen 2000 árboles, el número máximo de características a considerar en cada nodo, evaluado en un rango entre 2 y 7, y el tamaño mínimo de las hojas, probado entre valores de 1 y 11. La selección de estos hiperparámetros fue optimizada mediante validación cruzada, además, se utilizó el error out-of-bag para estimar el error de generalización. En lo que respecta a la evaluación del modelo, el conjunto de datos fue dividido en un conjunto de entrenamiento y uno de prueba para evaluar el rendimiento del modelo en datos no vistos.

A diferencia del modelo probit, el objetivo del modelo Random Forest no es estimar los efectos de x_j sobre $P(y = 1 | x)$, sino encontrar la mejor manera de categorizar la variable dependiente y con la información disponible en x . Sin embargo, Random Forest permite calcular una medida de importancia de las variables, basada en la reducción promedio de impureza en los nodos del bosque generada por cada variable independiente. Si bien estas medidas no permiten inferencias causales ni direccionales, ofrecen evidencia empírica sobre las variables que son más relevantes para la clasificación del empleo informal, y por ello, pueden considerarse como “posibles determinantes” dentro del marco del modelo.

Con el fin de aprovechar las ventajas de las TT econométricas y algunas características de los métodos de ML, se utiliza una técnica que combina componentes de ambas metodologías, el método Adaptive Lasso. Este puede rescatar la estructura econométrica tradicional, manteniendo un enfoque en la interpretabilidad y a su vez, mediante un enfoque basado en la validación cruzada (técnica comúnmente asociada con ML) se puede realizar la selección automática de relaciones no lineales entre las determinantes. Este es un método de regresión que mejora el método Lasso tradicional, introduce pesos adaptativos en la penalización asociada a los parámetros para lograr una selección consistente de variables y reducir el sobreajuste [55]. Esencialmente, este método es una secuencia de estimaciones Lasso con validación cruzada, donde cada aproximación Lasso es al menos tan parsimoniosa como la anterior. Primero se estiman los coeficientes $\hat{\beta}^{lasso}$, luego, se excluyen los coeficientes iguales a cero y se construye la matriz de pesos para las covariables distintas de cero $\hat{w}_j = \frac{1}{|\hat{\beta}^{lasso}|}$ y se estima $\hat{\beta}^{*(n)}$, adicionalmente para un modelo Adaptive Lasso con tres niveles (el utilizado en esta investigación) se vuelven a estimar los pesos con los coeficientes de la última estimación y se estima $\hat{\beta}^{*(n)}$ nuevamente.

$$2) \hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

$$3) \hat{\beta}^{*(n)} = \underset{\beta}{\operatorname{argmin}} \left\| y - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda_n \sum_{j=1}^p \hat{w}_j |\beta_j|$$

donde λ y λ_n controlan la fuerza de la penalización y se identifican mediante validación cruzada, a medida que dichos parámetros incrementan el valor óptimo de los coeficientes más pequeños es cero.

En el contexto de esta investigación para la identificación de las relaciones no lineales entre las determinantes se estima un modelo de probabilidad lineal Adaptive Lasso con tres niveles donde λ y λ_n se eligen mediante validación cruzada con 10 grupos. Como variables independientes se emplean las mismas utilizadas en el modelo probit, incluyendo las potencias hasta el tercer grado de las variables continuas, así como todas las interacciones posibles entre variables continuas y categóricas. Las variables de control no forman parte del proceso de selección penalizado, así que no se consideran las interacciones con estas, pero estas variables sin modificación se mantienen en todas las estimaciones Adaptive Lasso.

Tras identificar las variables relevantes mediante la metodología descrita, se estima un modelo probit que incorpora tanto los términos principales como las potencias e interacciones identificadas por el modelo Adaptive Lasso. Finalmente, se calculan los efectos marginales promedio para capturar los efectos de x_j sobre $P(y = 1 | x)$ considerando nuevas relaciones no lineales

entre las determinantes, no comúnmente identificadas en la literatura.

Por último, como parte de la evaluación de los modelos de clasificación, se analizarán las siguientes métricas de desempeño. La sensibilidad, representa la capacidad del modelo para identificar correctamente los casos positivos. La especificidad, representa la capacidad del modelo para identificar correctamente los casos negativos. El porcentaje de observaciones correctamente clasificadas ofrece una medida global de la exactitud predictiva del modelo. El área bajo la curva ROC proporciona una evaluación integral de la capacidad discriminativa del modelo al considerar todos los posibles umbrales de clasificación, un valor de 1 indica que se cuenta con discriminación perfecta de clases y 0.5 sugiere un rendimiento aleatorio. Estas métricas permitirán evaluar de manera comprehensiva el equilibrio entre los diferentes tipos de error de clasificación y el rendimiento general del modelo.

4. Resultados

Los resultados se presentan en el siguiente orden, primero, los efectos marginales promedio del modelo Probit tradicional (con los coeficientes completos en el Apéndice A); luego, las variables más relevantes identificadas mediante la reducción promedio de impureza del Random Forest; a continuación, los efectos marginales del modelo Probit extendido con términos no lineales elegidos mediante el método Adaptive Lasso (con los coeficientes completos en el Apéndice B); y finalmente, la comparación de las métricas de desempeño para todos los modelos estimados.

La Tabla 2 presenta los efectos marginales promedio estimados mediante el modelo Probit para los valores normalizados de las determinantes del empleo informal en Bolivia, utilizando datos de las Encuestas de Hogares (EH) 2022 y 2023. Los resultados muestran patrones consistentes en ambos años: variables como la edad y el género tienen efectos positivos y significativos sobre la probabilidad de informalidad, mientras que los años de estudio y los ingresos del hogar presentan efectos negativos. Destacan las diferencias territoriales, con mayor informalidad en áreas rurales y variaciones significativas por departamento (ej. efectos negativos en Santa Cruz y Beni). La pertenencia a grupos originarios y la cantidad de personas menores a 5 años en el hogar solo fue significativa en 2022.

Una perspectiva diferente se puede observar en los resultados del modelo Random Forest, donde las características que mejor ayudan a predecir el empleo informal son los ingresos laborales, los años de estudio, la edad y los ingresos del hogar per cápita, en ese orden, para las EH 2022 y EH 2023.

Tabla 2: Efectos marginales promedio – Modelo Probit, 2022-2023

Variable	EH 2022		EH 2023	
	Coef.	Err. Est.	Coef.	Err. Est.
Edad	0.0463***	(0.0042)	0.0520***	(0.0042)
Años de estudio	-0.1147***	(0.0045)	-0.1438***	(0.0049)
Ingreso del hogar por persona	-0.0725***	(0.0052)	-0.0503***	(0.0077)
Ingreso no laboral	0.0315***	(0.0054)	0.0269***	(0.0049)
Género (ref. Hombre)				
Mujer	0.0738***	(0.0073)	0.0691***	(0.0073)
Pertenencia a un grupo originario (ref. No pertenece)				
Pertenece	0.0183*	(0.0095)	0.0003	(0.0094)
Personas menores a 5 años	-0.0147***	(0.0041)	-0.0038	(0.0039)
Área (ref. urbana)				
Rural	0.1299***	(0.0106)	0.1364***	(0.0100)
Departamento (ref. La Paz)				
Chuquisaca	-0.0573***	(0.0163)	-0.0278*	(0.0151)
Cochabamba	0.0147	(0.0113)	-0.0119	(0.0110)
Oruro	-0.0023	(0.0147)	-0.0233	(0.0144)
Potosí	-0.0077	(0.0149)	-0.0355**	(0.0145)
Tarija	-0.0274*	(0.0149)	-0.0055	(0.0140)
Santa Cruz	-0.0554***	(0.0111)	-0.0483***	(0.0114)
Beni	-0.0543***	(0.0169)	-0.0428***	(0.0156)
Pando	-0.1005***	(0.0188)	-0.0049	(0.0161)
***p≤0.01; **p≤0.05; *p≤0.10				

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

Tabla 3: Reducción De Impureza Promedio – Random Forest, 2022-2023

Característica	EH 2022	EH 2023
Ingreso laboral	0.2408	0.2311
Años de estudio	0.1164	0.1243
Edad	0.0859	0.0828
Ingresos del hogar por persona	0.0590	0.0624

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

Es relevante reconocer que si bien mediante la perspectiva econométrica tradicional no se suele incluir al ingreso laboral como una determinante, principalmente por los problemas de causalidad reversa y endogeneidad relacionados a utilizarla como variable independiente, esta es la variable asociada con la mayor reducción promedio de las impurezas para ambas encuestas de hogares, reflejando el rol y abriendo la discusión a considerar a los ingresos laborales como una variable especialmente relevante al momento de comprender la decisión de ser un empleado informal. Asimismo, estos resultados refuerzan la importancia en términos de poder predictivo de los años de estudio, la edad y los ingresos del hogar per cápita.

Por otro lado, considerando los resultados del Adaptive Lasso, también se presentan los efectos marginales promedio de los modelos considerando las relaciones no lineales específicas para cada estimación, se observa que hay efectos marginales promedio que se mantiene relativamente iguales, edad, años de estudio, miembros en el hogar menores a cinco años; mientras los efectos de los ingresos del hogar per cápita, ingresos no laborales y pertenecer a una nación originaria varían notablemente.

Las relaciones identificadas mediante el Adaptive Lasso revelan patrones no lineales que ilustran la complejidad de los determinantes de la informalidad. La relación cuadrática en la edad y educación muestra cómo el efecto sobre la informalidad se intensifica progresivamente en grupos de mayor edad y con mayor educación, para los ingresos del hogar, la curva en U invertida revela que el efecto de los ingresos per cápita del hogar reducen a medida que dichos ingresos incrementan. Las interacciones significativas, como el efecto combinado entre género y educación, penaliza particularmente a las mujeres con mayor formación, y la relación entre ingresos per cápita del hogar e ingresos no laborales, destaca cómo estos factores no operan de forma aislada, sino mediante mecanismos interdependientes, con efectos diferenciados para la población originaria, donde la pertenencia a grupos indígenas interactúa con el género y los ingresos per cápita del hogar, configurando un panorama donde la informalidad emerge como resultado de múltiples dinámicas entrelazadas.

Finalmente, comparando las métricas de desempeño de los modelos estimados, se puede observar que el Random Forest supera consistentemente a los demás modelos en capacidad predictiva, mostrando el mejor equilibrio entre sensibilidad y especificidad, junto con un destacado poder discriminatorio (área debajo de la curva ROC cercana al 90%). El Probit, aunque efectivo para identificar casos de informalidad, presenta limitaciones para clasificar correctamente a los trabajadores formales, reflejando las restricciones de los modelos paramétricos ante fenómenos complejos. El Probit con Adaptive Lasso ocupa una posición intermedia, confirmando que los enfoques híbridos mejoran sobre las técnicas tradicionales, aunque aún sin alcanzar el rendimiento de ML. Estos patrones se mantienen estables entre 2022 y 2023, sugiriendo que las diferencias responden a ventajas metodológicas.

Los resultados revelan la naturaleza multifacética de los determinantes de la informalidad en Bolivia. El análisis Probit tradicional identifica patrones consistentes: efectos positivos de edad y género, e impactos negativos de educación e ingresos del hogar. El Random Forest, por su parte, destaca el rol predictivo central de los ingresos laborales, una variable tradicionalmente excluida por problemas de endogeneidad y valida la relevancia de los años de estudio, la edad y los ingresos del hogar per cápita. Los modelos híbridos (Probit con Adaptive Lasso) descubren relaciones no lineales clave (efectos cuadráticos en edad y educación, una curva en U invertida para ingresos del hogar, e interacciones significativas como género-educación, género-ingresos no laborales, género-originario) que afectan a los efectos marginales promedio de algunas variables. En términos de desempeño predictivo, el Random Forest supera consistentemente a los demás modelos, mientras el Probit muestra limitaciones en especificidad, y los enfoques híbridos ocupan una posición intermedia. Esta evidencia conjunta sugiere que la informalidad surge de dinámicas complejas, normalmente ignoradas en una aproximación basada únicamente en TT econométricas, se muestra que factores económicos, demográficos y territoriales interactúan de manera no lineal, requiriendo aproximaciones analíticas que combinen la capacidad explicativa de los modelos tradicionales con el poder predictivo de técnicas de ML para su mejor comprensión.

Tabla 4: Efectos marginales promedio – Modelo Probit con Adaptive Lasso, 2022-2023

Variable	EH 2022		EH 2023	
	Coef.	Err. Est.	Coef.	Err. Est.
Edad	0.0377***	(0.0002)	0.0572***	(0.0002)
Años de estudio	-0.1048***	(0.0002)	-0.1410***	(0.0002)
Ingreso del hogar por persona	-0.1288***	(0.0004)	-0.0839***	(0.0003)
Ingreso no laboral	0.0423***	(0.0003)	0.0243***	(0.0002)
Género (ref. Hombre)				
Mujer	0.0772***	(0.0004)	0.0757***	(0.0003)
Pertenencia a un grupo originario (ref. No pertenece)				
Pertenece	0.0142***	(0.0004)	-0.0052***	(0.0004)
Personas menores a 5 años	-0.0106***	(0.0002)	-0.0167***	(0.0002)
Área (ref. urbana)				
Rural	0.1111***	(0.0005)	0.1240***	(0.0004)
Departamento (ref. La Paz)				
Chuquisaca	-0.0346***	(0.0009)	-0.0023***	(0.0009)
Cochabamba	0.0261***	(0.0005)	0.0042***	(0.0005)
Oruro	-0.0034***	(0.0009)	-0.0093***	(0.0009)
Potosí	-0.0053***	(0.0008)	-0.0136***	(0.0008)
Tarija	-0.0119***	(0.0008)	0.0171***	(0.0008)
Santa Cruz	-0.0427***	(0.0005)	-0.0346***	(0.0005)
Beni	-0.0535***	(0.0010)	-0.0346***	(0.0009)
Pando	-0.0904***	(0.0016)	0.0097***	(0.0015)
***p≤0.01; **p≤0.05; *p≤0.10				

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

Tabla 5: Métricas de desempeño – Probit, Random Forest, Probit con Adaptive Lasso, 2022-2023

Indicador	EH 2022			EH 2023		
	Probit	Random Forest	Probit con Adaptive Lasso	Probit	Random Forest	Probit con Adaptive Lasso
Sensitividad	0.8510	0.8800	0.8516	0.8279	0.8900	0.8429
Especificidad	0.4794	0.6700	0.5011	0.5393	0.6500	0.5215
Observaciones correctamente clasificadas	0.7186	0.8072	0.7267	0.7240	0.8066	0.7272
Área bajo la curva ROC	0.7689	0.8800	0.7858	0.7803	0.8900	0.7917

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

5. Conclusiones y recomendaciones

Los resultados del modelo Probit identifican como determinantes significativas de la informalidad a la edad, años de estudio, ingresos del hogar per cápita e ingresos no laborales, mientras que pertenecer a un grupo originario o tener menores de cinco años en el hogar no resultan estadísticamente relevantes en este enfoque tradicional. Por su parte, el Random Forest resalta la importancia predictiva de los ingresos laborales (variable tradicionalmente omitida en el Probit) y valida la importancia de los años de estudio, edad e ingresos del hogar per cápita como potenciales determinantes del empleo informal dado su rol central en la clasificación del empleo informal. Finalmente, el Adaptive Lasso revela que las determinantes tradicionales deben analizarse desde un enfoque no lineal, ya que este método al capturar interacciones complejas y potencias de las variables modifica la significancia de variables que pueden ser ignoradas desde una perspectiva tradicional (como el origen étnico y la presencia de niños pequeños). Estos hallazgos subrayan que la comprensión integral del empleo informal exige evaluar no solo variables aisladas, sino también relaciones no lineales y contextos específicos que los modelos lineales tradicionales pasan por alto a su vez que se necesita profundizar en las estructuras teóricas que puedan vincular los ingresos laborales con el empleo informal explícitamente.

Estos hallazgos sugieren diseñar políticas diferenciadas que consideren tanto los determinantes tradicionales como las complejas interacciones identificadas. Para las variables significativas en el modelo Probit (edad, educación e ingresos), se recomiendan programas de capacitación laboral enfocadas en los adultos mayores y programas que favorezcan el logro educativo. Los resultados del Random Forest destacan la necesidad de incorporar los ingresos laborales en el diseño de políticas, proponiendo esquemas de formalización progresiva que ofrezcan beneficios fiscales y de seguridad social adaptados a distintos niveles de ingresos. El análisis no lineal del Adaptive Lasso revela la importancia de políticas con enfoque interseccional, particularmente para grupos originarios y hogares con niños pequeños. La integración de estas aproximaciones metodológicas en el ciclo de políticas públicas puede permitir el diseño de intervenciones más precisas y adaptadas a la compleja realidad del empleo informal en Bolivia.

Para futuras investigaciones, se recomienda, aprovechar la disponibilidad de la Encuesta Continua de Empleo (ECE), ya que incluye variables clave como tipo de contrato y registro tributario que ofrecen la posibilidad de una identificación más precisa del empleo informal, así como utilizar técnicas avanzadas dada su naturaleza de datos rotatorios de panel. Asimismo, futuras investigaciones pueden incorporar modelos como redes neu-

ronales y máquinas de vectores de soporte si se busca mejorar la clasificación de los empleados formales e informales. Asimismo, estos métodos predictivos podrían combinarse con técnicas de inferencia causal (Causal Forests, Double/debiased ML) para evaluar impactos de políticas. La implementación de estos enfoques es viable con los datos existentes y permitiría superar limitaciones de los modelos tradicionales, generando evidencia más robusta sobre los determinantes y dinámicas del empleo informal.

En resumen, el estudio revela que el empleo informal en Bolivia responde a factores complejos, el modelo tradicional Probit confirma el impacto de edad, educación e ingresos; el Random Forest destaca el rol clave de los ingresos laborales; y el modelo Probit con Adaptive Lasso descubre interacciones críticas y no linealidades que se traducen en resaltar la importancia de variables como la pertenencia a grupos originarios y contar con niños pequeños en el hogar. Estos hallazgos demuestran que combinar técnicas econométricas tradicionales con machine learning permite una comprensión más completa del fenómeno, esencial para diseñar políticas efectivas.

Referencias

- [1] Canelas, C. y Niño-Zarazúa, M. (2022). Informality and pension reforms in Bolivia: The case of Renta Dignidad. *The Journal of Development Studies*, 58 (7), 1436-1458. <https://doi.org/10.1080/00220388.2022.2032678>
- [2] Villalpando, M. N. (2024). Explorando las raíces de la informalidad laboral en Bolivia: Un análisis de sus determinantes. *ARU Search*, 5 (1), 181-207.
- [3] Banco Mundial. (2008). Aportes a una nueva visión de la informalidad laboral en la Argentina. Ministerio de Trabajo, Empleo y Seguridad Social.
- [4] Desai, A. (2023). Machine learning for economics research: When, what and how? *arXiv*. <https://arxiv.org/abs/2301.05026>
- [5] Zheng, E., Tan, Y., Goes, P., Chellappa, R., Wu, D., Shaw, M. y Sheng, O. (2017). When econometrics meets machine learning. *Data and Information Management*, 1 (2), 75-83. <https://doi.org/10.1515/dim-2017-0009>
- [6] Athey, S. e Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685- 725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- [7] Hussein, A.-O. (2020). How machine learning affects economics/econometrics? A critical review of machine learning and econometrics. Oxford Brookes Business School.
- [8] Pérez-Pons, M. E., Parra-Dominguez, J., Omatu, S., Herrera-Viedma, E. y Corchado, J. M. (2022). Machine learning and traditional econometric models: A systematic mapping study. *Journal of Artificial Intelligence and Soft Computing Re-*

- search, 12 (1), 5-20. <https://doi.org/10.2478/jaiscr-2022-0001>
- [9] Mullainathan, S. y Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31 (2), 87-106. <https://doi.org/10.1257/jep.31.2.87>
- [10] Yenilmez Oğuz, S. (2024). Machine learning integration in econometric models. *Next Generation Journal for the Young Researchers*, 8 (1), 77-80.
- [11] Ramos, 2024. [Referencia no encontrada en la bibliografía original].
- [12] Dell'Anno, R. (2021). Theories and definitions of the informal economy: A survey. *Journal of Economic Surveys*, 35 (5), 1610-1643. <https://doi.org/10.1111/joes.12450>
- [13] Organización Internacional del Trabajo (OIT). (2015). Policy responses to the informal economy. Oficina Internacional del Trabajo.
- [14] Grusky, D. B. (2008). Social stratification. Westview Press.
- [15] Losby, J. L., Else, J. F., Kingslow, M. E., Edgcomb, E. L., Malm, E. T. y Kao, V. (2002). Informal economy literature review. ISED Consulting and Research.
- [16] Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics*, 12, 525-546. <https://doi.org/10.1146/annurev-economics-082119-121914>
- [17] Freije, S. (2002). Informal employment in Latin America and the Caribbean: Causes, consequences and policy recommendations. Banco Interamericano de Desarrollo.
- [18] Thomas, R. y Thomas, H. (1994). The informal economy and local economic development policy. *Local Government Studies*, 20 (3), 486-501. <https://doi.org/10.1080/03003939408433740>
- [19] Ulyssea, G., Bobba, M. y Gadenne, L. (2023). Informality. *VoxDevLit*. <https://voxdev.org>
- [20] Organización Internacional del Trabajo (OIT). (2023). Statistics on the informal economy. <https://www.ilo.org/informaleconomy>
- [21] Román, S. (2011). Costos laborales, economía informal y reformas a la legislación laboral en Bolivia.
- [22] Miranda Gonzales, Y. (2018). Estimación de la Economía Informal en Bolivia. *Revista Varianza*, 15-29.
- [23] Jemio, V. (2022). El Rol del Sector Informal en Bolivia: Una Primera Mirada a Través de un Enfoque de Equilibrio General Dinámico Estocástico. *Cuadernos de Investigación Económica Boliviana*, 5(2), 65-116.
- [24] Unidad de Análisis de Políticas Sociales y Económicas (2022). Diagnóstico del Empleo, Sector y Economía Informal en Bolivia. La Paz.
- [25] Cimoli, M., Primi, A., & Pugno, M. (2006). Un modelo de bajo crecimiento: la informalidad como restricción estructural. *Revista de la CEPAL*, 88, 89-107.
- [26] Morales, R. (2008). El sector informal en Bolivia. Reflexiones teóricas y realidad estadística. La Paz, Bolivia: Superintendencia de Empresas.
- [27] Muriel, B. (2018). EL Círculo Vicioso entre Elevada Informalidad y Baja Productividad en Bolivia. La Paz, Bolivia: Agencia Inteligente para el Desarrollo de Bolivia.
- [28] Velasco Valdez, D. (2015). Economía informal en Bolivia: análisis, evaluación y cuantificación en base al enfoque monetario de la demanda de efectivo (periodo 1994-2014). *Investigación & Desarrollo*, 2(15), 76-89.
- [29] Molina Fernández, J. M. (2019). The dynamics of labor informality and international trade in Bolivia. *Economía Coyuntural*, 4(3), 11-36.
- [30] Novella, R., & Valencia, H. (2022). Active Labor Market Policies in a Context of High Informality: The Effect of PAE in Bolivia. *The Journal of Development Studies*, 2583-2603.
- [31] Angel-Urdinola, D. F. y Tanabe, K. (2012). Micro-determinants of informal employment. En *Striving for better jobs: The challenge of informality in the Middle East and North Africa region*.
- [32] Cerquera-Losada, O. H., Arias-Barrera, C. J. y Rojas-Velásquez, L. (2020). Determinantes del subempleo en Colombia: Una aproximación a partir de un modelo PROBIT. *El Ágora USB*, 20 (1), 157-172. <https://doi.org/10.21500/16578031.4525>
- [33] Coaquira-Velásquez, M. A. (2021). Determinantes de la informalidad y subempleo en el departamento de Puno: Un modelo Probit bivariado aplicado para el año 2019. *Semestre Económico*, 10 (1), 49-67.
- [34] Duval-Hernández, R. (2022). Choices and constraints: The nature of informal employment in urban Mexico. *Journal of Development Studies*, 58 (7), 1349-1362. <https://doi.org/10.1080/00220388.2022.2032677>
- [35] Espinoza, J. R. (2021). Análisis de la informalidad comercial en el distrito de San Juan de Miraflores. *Horizonte Empresarial*, 20 (2), 782-792.
- [36] Liang, Z., Appleton, S. y Song, L. (2016). Informal employment in China: Trends, patterns and determinants of entry. *IZA Discussion Paper*, N.º 10139.
- [37] Muhammad Tanveer, A. K. y Hussain, B. (2021). Measurement and determinants of informal employment: Evidence from Pakistan. *Pakistan Social Science Review*, 5 (2), 309-324.
- [38] Narváez, A. R., López, J. W. y Ochoa, C. L. (2021). Causas de la informalidad laboral en Montería, Colombia: Un modelo econométrico Probit. *Revista de Economía y Finanzas*, 15 (44), 45-60.
- [39] Onofrei, N. (2024). The determinants of informal economy in Eastern European countries. *Economy and Sociology*, 2, 45-58.
- [40] Ortiz, C. H., Uribe, J. I. y García, G. A. (2007). Informalidad y subempleo: Un modelo probit bivariado aplicado al Valle del Cauca. *Sociedad y Economía*, 13, 104-131.
- [41] Ticona-Carrizales, L., Portillo, H. A., Rodríguez-Limachi, O. M., Incacutipa-Limachi, D. J., Yapuchura-Saico, C. R., Huayta-Flores, L. y Ticona-Campos, V. N. (2024). Socioeconomic determinants of labor informality in the Department of Puno, Peru. *Journal of Ecohumanism*, 3 (8), 7235-7247. <https://doi.org/10.33140/JEH.03.08.02>

- [42] Andersen, L., & Muriel, B. (2007). Informality and Productivity in Bolivia: A Gender Differentiated Empirical Analysis. Development Research Working Paper Series, No. 07/2007, Institute for Advanced Development Studies (INESAD).
- [43] Tannuri-Pianto, M., Pianto, D., & Arias, O. (2004). Informal Employment in Bolivia: A lost proposition? World Bank mimeo.
- [44] Mendieta, P. (2022). Años perdidos: informalidad y retornos a la educación en Bolivia. *Revista Latinoamericana de Desarrollo Económico*, (37), 87-111.
- [45] Urquidi, M., Valencia, H., & Durand, G. (2021). Brecha de ingresos laborales por género en Bolivia. Un análisis de su evolución en el periodo 1993 a 2018. *Revista de análisis económico*, 36(2), 95-124.
- [46] López, E., Magalhães, L., Araya, I., & Guizardi, M. (2023). La provisión De Cuidados En Bolivia: Desigualdades De género Y Sobrecarga Femenina. *Revista Austral De Ciencias Sociales*, (45), 41-58.
- [47] Muriel, B., & Jemio, L. (2010). Mercado laboral y reformas en Bolivia. Development Research Working Paper Series, No. 07/2010.
- [48] Claure, M., Leyton, A., Valencia, C., Sánchez, K., & Davalos, J. (2017). Evidence of the Impacts of Minimum Wages on Labor Market Outcomes: The Case of Bolivia. Partnership for Economic Policy Working Paper.
- [49] Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7^a ed.). Cengage Learning.
- [50] Hastie, T., Tibshirani, R. y Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2^a ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- [51] Agrawal, A. (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.
- [52] Steuer, F. (2018). *Machine learning for public policy making: How to use data driven predictive modeling for the social good*. Student Paper Series. Hertie School.
- [53] Athey, S. y Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies*, 5 (1), 37-51.
- [54] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. y Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21 (1), C1-C68. <https://doi.org/10.1111/ectj.12097>
- [55] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101 (476), 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [56] Instituto Nacional de Estadística (INE). (2022). Bolivia - Encuesta de Hogares 2022 (EH 2022). <https://anda.ine.gob.bo/index.php/catalog/106>
- [57] Instituto Nacional de Estadística (INE). (2023). Bolivia - Encuesta de Hogares 2023 (EH 2023). <https://anda.ine.gob.bo/index.php/catalog/108>
- [58] Breiman, L. (2001). Random forests. *Machine Learning*, 45 (1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Apéndices

Apéndice A: Parámetros Estimados – Modelo Probit

Variable	EH 2022		EH 2023	
	Coef.	Err. Est.	Coef.	Err. Est.
Edad	0.1528***	(0.0142)	0.1756***	(0.0007)
Años de estudio				
Ingreso del hogar por persona	0.2447***	(0.0244)	0.2341***	(0.0012)
Ingreso no laboral	-0.3789***	(0.0158)	-0.4858***	(0.0008)
Género (ref. Hombre)	-0.2394***	(0.0177)	-0.1697***	(0.0007)
Mujer	-0.0484***	(0.0136)	-0.0127***	(0.0006)
Pertenencia a un grupo originario (ref. No pertenece)				
Pertenece	0.0606*	(0.0315)	0.0012	(0.0015)
Personas menores a 5 años	0.1041***	(0.0180)	0.0909***	(0.0007)
Área (ref. urbana)				
Rural	0.4284***	(0.0367)	0.4585***	(0.0015)
Departamento (ref. La Paz)				
Chuquisaca	-0.1866***	(0.0524)	-0.0943***	(0.0030)
Cochabamba	0.0499	(0.0383)	-0.0407***	(0.0019)
Oruro	-0.0077	(0.0492)	-0.0792***	(0.0030)
Potosí	-0.0256	(0.0496)	-0.1199***	(0.0026)
Tarija	-0.0905*	(0.0488)	-0.0190***	(0.0029)
Santa Cruz	-0.1805***	(0.0360)	-0.1625***	(0.0016)
Beni	-0.1770***	(0.0543)	-0.1444***	(0.0032)
Pando	-0.3220***	(0.0592)	-0.0169***	(0.0052)
Constante	0.2850***	(0.0289)	0.3232***	(0.0015)
Observaciones	19032		19066	
Wald-Chi²	0.0000		0.0000	
Pseudo-R²	0.1716		0.1911	
*** p≤0.01; ** p≤0.05; * p≤0.10				

Fuente: Elaboración propia en base a la EH 2022 y EH 2023

Apéndice B: Parámetros Estimados – Modelo Probit con Adaptive Lasso

Variable	EH 2022		EH 2023	
	Coef.	Err. Est.	Coef.	Err. Est.
Edad	0.3133***	(0.0014)	0.2318***	(0.0171)
Edad ²	0.0996***	(0.0007)	0.0756***	(0.0136)
Edad ³	-0.0538***	(0.0005)	-	-

Años de estudio	-0.2038***	(0.0010)	-0.3311***	(0.0215)
Años de estudio ²	-0.0995***	(0.0007)	-0.1309***	(0.0155)
Ingresos del hogar por persona	-0.3655***	(0.0014)	-0.2199***	(0.0294)
Ingreso del hogar por persona ²	0.1810***	(0.0008)	0.0111	(0.0018)
Ingreso del hogar por persona ³	-0.0141***	(0.0001)	-	-
Ingreso no laboral	0.1490***	(0.0012)	0.0988***	(0.0211)
Género (Mujer)	0.2961***	(0.0015)	0.2693***	(0.0312)
Género y Etnia (Hombre y Pertenece)	0.0329***	(0.0019)	-0.0913**	(0.0398)
Género y Etnia (Mujer y Pertenece)	0.0918***	(0.0025)	0.1069**	(0.0511)
Género y Edad (Mujer)	-0.0455***	(0.0015)	-	-
Etnia y Edad (pertenece)	-0.001	(0.0015)	-0.0151	(0.0327)
Género y Años de estudio (Mujer)	-0.3053***	(0.0015)	-0.2761***	(0.0300)
Etnia e Ingreso del hogar por persona (Pertenece)	-0.1766***	(0.0017)	-0.2429***	(0.0416)
Género y Personas menores a 5 años (Hombre)	-0.0529***	(0.0009)	-0.1015***	(0.0213)
Género y Personas menores a 5 años (Mujer)	-0.0167***	(0.0011)	-0.0055	(0.0235)
Años de estudio e Ingreso del hogar por persona	-0.1503***	(0.0010)	-0.0267	(0.0176)
Años de estudio e Ingreso no laboral	-0.0148***	(0.0009)	-	-
Ingreso del hogar por persona y Personas menores a 5 años	0.0657***	(0.0010)	-0.1229***	(0.0301)
Ingreso del hogar por persona e Ingreso no laboral	-0.0226***	(0.0003)	-	-
Género e Ingreso no laboral (Mujer)	-	-	-0.0377	(0.0278)
Edad y Años de estudio	-	-	0.0167	(0.0155)
Años de estudio y Personas menores a 5 años	-	-	0.0545***	(0.0166)
Área (Rural)	0.3796***	(0.0016)	0.4302***	(0.0362)
Departamento (Cochabamba)	-0.1174***	(0.0030)	-0.0081	(0.0519)
Departamento (Oruro)	0.0918***	(0.0019)	0.0148	(0.0381)
Departamento (Potosí)	-0.0118***	(0.0031)	-0.0327	(0.0497)
Departamento (Tarija)	-0.0184***	(0.0027)	-0.0476	(0.0497)
Departamento (Santa Cruz)	-0.0409***	(0.0029)	0.0606	(0.0488)
Departamento (Beni)	-0.1444***	(0.0017)	-0.1198***	(0.0383)
Departamento (Pando)	-0.1804***	(0.0033)	-0.1199**	(0.0525)
Departamento (Cochabamba)	-0.3004***	(0.0053)	0.0343	(0.0559)
Constante	0.2071***	(0.0018)	0.3011***	(0.0363)
Observaciones	19032		19066	
Wald-Chi ²	0.0000		0.0000	
Pseudo-R ²	0.1995		0.2136	
***p≤0.01; **p≤0.05; *p≤0.10				

Fuente: Elaboración propia en base a la EH 2022 y EH 2023