PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INFRAESTRUCTURA HABITAD Y CREATIVIDAD

MAESTRÍA EN SISTEMAS DE INFORMACIÓN MENCIÓN DATA SCIENCE



Trabajo De Titulación

Tema:

El costo de oportunidad de la informalidad laboral: una estimación de la penalización salarial y perfiles de riesgo asociados a ella en Ecuador

Autor

Oscar Leonardo Enríquez Sánchez

Director:

Henry Nelson Roa Marin, PhD.

Quito, agosto de 2026

Al costo de oportunidad que supuso la decisión de seguir aquí sin levantarme hasta finalizar, y a la constante renuncia a tomar el camino fácil. A las batallas por lo pequeño que, aunque sabias parecían innecesarias e imprudentes; a los adversarios ganados, a los amigos que quedaron atrás, pero sobre todo a quienes conociendo mis errores eligieron estar. A Dios, a mamá, a papá, a Fer, a Ruffo y a Raquel.

Resumen

Continuando con el esfuerzo de Enríquez (2014), esta investigación aborda la informalidad laboral en Ecuador combinando econometría convencional con técnicas de ciencia de datos, a partir de los microdatos de la ENEMDU-INEC (2025). Bajo un enfoque legalista de medición, el estudio dimensiona y caracteriza la informalidad, evidenciando su heterogeneidad según tipología ocupacional, nivel educativo, territorio, género y ciclo de vida, para luego estimar la penalización salarial neta asociada a esta condición mediante métodos paramétricos independientes que convergen en sus resultados.

Finalmente, el estudio construye perfiles de riesgo de informalidad mediante modelos de aprendizaje automático, identificando los determinantes que anteceden y acompañan la inserción laboral precaria, los cuales se sintetizan en una matriz de perfiles representativos. En conjunto, los resultados muestran que la informalidad ecuatoriana no es aleatoria en su origen ni gratuita en su ejercicio, y que su combate exige atender simultáneamente tanto el origen de esa condición como la forma desigual en que el mercado retribuye a formales e informales.

Índice de Contenidos

Capítulo 1: Introducción....7
Justificación:....7
Planteamiento del Problema:....9
Capítulo 2: Marco Teórico....11
Capítulo 3: Metodología:....22
Capítulo 4: Resultados....31
4.1. Dimensionamiento y Caracterización de la Informalidad Laboral en Ecuador....31
4.1.1. La magnitud del fenómeno: casi uno de cada dos trabajadores....31
4.1.2. Un país de dos velocidades: la brecha urbano-rural....34
4.1.3. Género e informalidad: un patrón más matizado de lo esperado....35
4.1.4. La heterogeneidad sectorial de la informalidad....37
4.1.5. El ciclo vital de la informalidad: una trayectoria ascendente tras la juventud....38
4.1.6. La educación como escudo: el gradiente educativo inverso....40
4.1.7. La brecha de ingresos: una primera aproximación monetaria....41
4.1.8. La brecha no es uniforme: penalización salarial por categoría ocupacional....43
4.1.9. Trayectorias de ingreso: dualismo estructural y brechas de género....44
4.1.10. Síntesis: de la caracterización a la estimación....45
4.2. Estimación de la Penalización Salarial Asociada a la Informalidad Laboral....46
4.2.1. El punto de partida: la insuficiencia de la ecuación de Mincer clásica....47
4.2.2. La ecuación extendida: hacia una lectura estructural del ingreso laboral....52
4.2.3. Validación desde un enfoque alternativo: la descomposición de Oaxaca-Blinder....59
4.3. Determinación de Perfiles de Riesgo de Informalidad Laboral....70
Regresión logística: planteamiento y determinantes....70
Gradient Boosting: planteamiento y determinantes....74
De los coeficientes a la explicación individual: por qué recurrir a LIME....76
4.3.2. Lo que el Origen Social Determina: los Márgenes de Acción de la Política Pública....77
4.3.3. La Decisión que Parece Rentable y no lo Es....78
4.3.4. Grupos Invisibles para la Política de Combate a la Informalidad....79
4.3.5. Síntesis: la Matriz de Perfiles de Riesgo....81
Conclusiones....87
Recomendaciones....93
Anexos:....98
Referencias Bibliográficas....94

Índice de Tablas

Cuadro 1. Diseño muestral. Informalidad en la ENEMDU 2025 25
Cuadro 2. Composición de la informalidad laboral (valores ponderados por fexp) 33
Cuadro 3. Distribución de la población ocupada por área y categoría laboral 35
Cuadro 4. Tasa de informalidad laboral por género y área 36
Cuadro 5. Tasa de informalidad laboral por grupo de edad (quinquenal) 40
Cuadro 6. Tasa de informalidad laboral por nivel de instrucción 41
Cuadro 7. Ingreso laboral promedio, formal vs. informal 42
Cuadro 8. Ecuación de Mincer clásica (conjunto de entrenamiento) 49
Cuadro 9. Ecuación de Mincer extendida (conjunto de entrenamiento) 53
Cuadro 10. Regresión de Mincer extendida en el sector formal (conjunto de entrenamiento) 60
Cuadro 11. Regresión de Mincer extendida, sector informal (conjunto de entrenamiento) 63
Cuadro 12. Descomposición de Oaxaca-Blinder de la brecha salarial formal-informal 66
Cuadro 13. Comparación de coeficientes entre la ecuación conjunta y las regresiones formal e informal separadas (en porcentaje) 69
Cuadro 14. Regresión logística: determinantes del riesgo de informalidad (conjunto de entrenamiento). 70
Cuadro 15. Desempeño predictivo de la regresión logística 73
Cuadro 16. Desempeño predictivo del Gradient Boosting 74
Cuadro 17. Perfiles simulados: origen socioeconómico, territorial e intergeneracional 78
Cuadro 18. Perfiles simulados: decisión ocupacional y su costo esperado 79
Cuadro 19. Composición del hogar según el quintil de ingreso per cápita 80

Índice de Figuras

Figura 1. Línea de tiempo del estudio de la informalidad laboral (1949–1989). Fuente: elaboración propia con base en el recorrido histórico desarrollado en este capítulo 12
Figura 2. Línea de tiempo del estudio de la informalidad laboral (1990–2026). Fuente: elaboración propia con base en el recorrido histórico desarrollado en este capítulo 13
Figura 3. Extensión y delimitación de la informalidad 20
Figura 4. Composición de la informalidad laboral, ponderada por fexp 32
Figura 5. Tasa de informalidad laboral total (informal + servicio doméstico) por área 34

Figura 6. Tasa de informalidad laboral por género y área.....36
Figura 7. Tasa de informalidad laboral por grupo de edad.....39
Figura 8. Tasa de informalidad laboral por nivel de instrucción.....40
Figura 9. Distribución del logaritmo del ingreso laboral, formal vs. informal.....42
Figura 10. Ingreso laboral promedio ponderado, formal vs. informal, por categoría de ocupación 43
Figura 11. Promedio ponderado del logaritmo del ingreso laboral por grupo de edad, género y condición de formalidad.....45
Figura 12. Distribución del logaritmo del ingreso laboral, observado y predicho, formal vs. informal. 58
Figura 13. Convergencia de la penalización salarial por informalidad entre métodos independientes..74
Figura 14. Curvas ROC de la regresión logística y el Gradient Boosting (conjunto de prueba).....68
Figura 14. Curvas ROC de la regresión logística y el Gradient Boosting (conjunto de prueba).....75
Figura 15. Comparación de importancia de variables: regresión logística vs. Gradient Boosting.....76
Figura 16. Matriz de perfiles de riesgo de informalidad y penalización salarial..... 82

Capítulo 1: Introducción

Justificación:

La informalidad laboral es un problema engañosamente simple, pero también ampliamente complejo. A diario, las transacciones informales forman parte del paisaje urbano y rural, legitimadas por la ventaja de la flexibilización o de la economización monetaria inmediata que brinda la economía de la sombra en el corto plazo; siendo esta una realidad que, con el paso del tiempo, se ha empezado a aceptar y a la que la población paulatinamente se acostumbra sin dimensionar el costo de oportunidad al que se enfrenta. En este sentido, la informalidad, lejos de ser un fenómeno transitorio o temporal, se presenta como un fenómeno de profundos orígenes e implicaciones estructurales que moldea el desempeño de la economía y afecta directamente a todos los agentes.

Por una parte, la informalidad lejos de representar una alternativa ante el desempleo o de constituir un soporte frente a los desequilibrios del mercado laboral, en realidad es un fenómeno que relega, margina, e impide el desarrollo económico de los agentes a los que absorbe; pues al interior de la informalidad se generan factores de "auto-refuerzo" , tales como ausencia de información, la escaza estabilidad y falta de transparencia, que imposibilitan que desde la política pública se logre contrarrestar las condiciones de precariedad, de baja productividad e incluso de pobreza aquí existentes. En este sentido, quienes son absorbidos por la informalidad o deciden dar el paso hacia ella no contribuyen ni se benefician de la redistribución del ingreso, al mantenerse relegados e invisibles para la política pública. Así mismo, alimentan una economía de supervivencia que no transparenta sus costos operativos reales y legales, y se acostumbran a operar por mutuo acuerdo o por imposición de una de las partes al margen de la política pública y sus regulaciones.

Por otra parte, la informalidad tiene implicaciones que trascienden al propio sector y afectan a la sociedad en su conjunto, pues la provisión de bienes y servicios públicos recae y es producto de la contribución de la economía formal, de modo que cuando la informalidad gana terreno, como lo ha venido haciendo en los últimos años en CEPAL (2024), la presión fiscal crece sobre los contribuyentes, siendo éste un nuevo costo que incentiva a la informalidad por partida doble: quienes deciden permanecer en la formalidad reciben mayor presión fiscal y perciben que paulatinamente no existe retorno claro de su contribución. Por ello, la formalidad deja de ser asumida desde la perspectiva del incentivo y la responsabilidad ciudadana para ser interpretada como la obligación derivada del miedo a la regulación y al control. En esta perspectiva, la informalidad no sólo es una consecuencia de la precariedad, la falta de educación o la pobreza de los contribuyentes, sino que también representa un "modus operandi" para reducir artificialmente los costos operativos de quienes han aprendido a permanecer invisibles a la política pública y se sienten cómodos o menos aprisionados sin ella.

Como contrapartida de esta decisión y de la dinámica laboral y productiva, en la economía se genera un problema de free rider que debilita la institucionalidad nacional y resta eficiencia y calidad a la prestación de los servicios públicos.

Como consecuencia, la informalidad moldea el desarrollo de la economía, a la vez que marca su ritmo de crecimiento, pues la política pública no puede incidir en la totalidad de la base productiva para generar incentivos, contribuir a su estabilidad e impulsar su desarrollo. Ello, sin hablar de las externalidades negativas hacia los grupos más vulnerables, derivadas de una menor recaudación fiscal, lo que

imposibilita que el Estado atienda a la primera infancia, a adultos mayores en situación de pobreza, a personas con discapacidad o a quienes hayan sufrido una vulneración de derechos. En este mismo sentido, el comercio justo pierde terreno y la competencia desleal empieza a convertirse en la regla y modus operandi del emprendimiento ¿Cómo puede competir en costos el comerciante que afilia a la seguridad social a sus trabajadores y que declara sus impuestos y mantiene al día sus licencias de operación, con el comerciante que decide operar con una masa laboral que gana menos del salario básico, que no cuenta con afiliación a la seguridad social y que no paga impuestos por su actividad económica?

Como puede verse, el fenómeno de la informalidad conlleva grandes externalidades negativas y marca el rumbo del desarrollo de un país. En consecuencia, se han desarrollado numerosos estudios sobre esta temática, acompañados de diversos intentos de describir, medir y estimar la magnitud de la informalidad laboral.

Sin embargo, actualmente la ciencia de datos pone a disposición de la humanidad una diversidad de técnicas para estimar realmente cuál es la penalización salarial asociada a la condición de informalidad laboral de los trabajadores, así como técnicas de aprendizaje automático para identificar cuáles son las características que mayoritariamente inciden en la probabilidad de ser informal. Aprovechar estas técnicas disponibles, para el estudio de la informalidad otorga un relevo fresco y necesario a la econometría convencional y a la economía laboral que permite apostar al entendimiento de esta problemática de la mano de insights robustos.

En este sentido, se consideran las siguientes preguntas relevantes: ¿Cuál es la penalización salarial asociada a la informalidad laboral en Ecuador y cómo varía entre distintos perfiles sociodemográficos, estimada mediante la ecuación de Mincer y los modelos salariales? ¿Qué características estructurales y coyunturales determinan la probabilidad de ser informal en la población económicamente activa del Ecuador? ¿Qué modelo de regresión ofrece la mayor precisión para estimar el costo de oportunidad de la informalidad en términos de ingresos laborales? ¿Existen diferencias significativas en la penalización salarial entre perfiles de riesgo definidos por género, edad, nivel educativo y tipo de ocupación? En consecuencia, este estudio busca responder estas importantes interrogantes y abordar este desafío de política pública mediante la aplicación de técnicas de Data Science que integran la ecuación de Mincer para la definición de brechas salariales con modelos avanzados de aprendizaje automático (Gradient Boosting) y la identificación de perfiles de riesgo, aportando rigor empírico y predictivo al análisis.

Como resultado, se espera que el estudio resulte relevante en distintos ámbitos. Puntualmente, su relevancia en el ámbito social radica en que la informalidad afecta directamente la estabilidad laboral, la equidad salarial y la sostenibilidad fiscal. Desde el punto de vista científico, la investigación contribuye a la literatura sobre ciencia de datos aplicada a la economía laboral al combinar enfoques econométricos clásicos con métodos de aprendizaje automático, lo que fortalece la capacidad para analizar fenómenos complejos. En el ámbito laboral y de la política pública, los resultados permitirán diseñar estrategias de formalización basadas en evidencia, optimizar la asignación de recursos y mejorar la eficiencia del gasto público. En este sentido, esta investigación no solo describe la informalidad, sino que también identifica sus desafíos y los traduce en métricas verificables que orientan las decisiones estratégicas en materia laboral.

Planteamiento del Problema:

La informalidad laboral en Ecuador constituye un fenómeno estructural que, lejos de ser transitorio, moldea el desempeño económico y limita la capacidad redistributiva del Estado. Tal como se expone en *El reto de la política fiscal en la informalidad* (Enríquez, 2014), la informalidad no solo representa una respuesta adaptativa ante el desempleo, sino también un entramado de relaciones económicas y sociales que perpetúan la precariedad, la baja productividad y la exclusión fiscal. En ese sentido, el problema no radica únicamente en la magnitud del sector informal, sino en la ausencia de instrumentos analíticos capaces de cuantificar con precisión el costo de oportunidad de ser informal, tanto para los individuos como para el Estado.

A pesar de los avances regionales en el estudio de la informalidad —por parte del CEDLAS, el Banco Mundial y el BID—, el caso ecuatoriano carece de investigaciones recientes que integren enfoques econométricos clásicos con técnicas de *machine learning*. Los estudios previos se han limitado a caracterizar la informalidad desde una perspectiva descriptiva o fiscal, sin estimar de manera rigurosa la penalización salarial ni los perfiles de riesgo asociados, ni el costo de oportunidad que supone para la población económicamente activa este fenómeno.

Por ello, un desafío relevante de política pública radica en estimar y comparar el costo de oportunidad y penalización laboral en términos salariales derivados de la informalidad laboral en Ecuador. Para ello, una alternativa útil, relevante y sofisticada es estimar el costo de oportunidad o la penalización salarial asociados a la informalidad, utilizando la Ecuación de Mincer como modelo teórico base y contrastando esta ecuación con algoritmos avanzados de regresión, tales como Random Forest o Gradient Boosting, de modo que permitan capturar relaciones no lineales entre variables socioeconómicas y este fenómeno. Comparar los resultados de estos modelos permitirá seleccionar la mejor opción para obtener estimaciones precisas.

Preguntas de Investigación:

Pregunta principal

¿Cuál es la penalización salarial asociada a la informalidad laboral en Ecuador y cómo varía entre distintos perfiles sociodemográficos, estimada mediante la ecuación de Mincer y modelos avanzados de machine learning?

Preguntas secundarias

1. ¿Qué características estructurales y coyunturales están asociadas a la probabilidad de ser informal en la población económicamente activa del Ecuador?

2. ¿Qué modelo de regresión (Mincer, Oaxaca Blinder) ofrece la mayor precisión para estimar el costo de oportunidad de la informalidad en términos de ingreso laboral?

3. ¿Qué perfiles socioeconómicos tienen mayor riesgo o probabilidad de caer en la informalidad laboral?

Objetivos General y Específicos:

Objetivo General:

Estimar el costo de oportunidad de la informalidad laboral en Ecuador a través de la penalización salarial asociada a esta mediante la aplicación de la ecuación de Mincer y Oaxaca-Blinder, empleando como fuente de información la Encuesta Nacional de Empleo y Desempleo (ENEMDU 2025) y determinar el riesgo de que los trabajadores caigan en la informalidad en función de sus distintos perfiles socioeconómicos.

Objetivos Específicos:

Objetivo específico 1.- Dimensionar y caracterizar la informalidad laboral en el Ecuador en términos de su magnitud respecto de la población económicamente activa (PEA) y su distribución según tipología ocupacional, género, área geográfica, nivel educativo y ciclo de vida, mediante un análisis exploratorio de datos con la ENEMDU 2025.

Objetivo específico 2.- Estimar la penalización salarial asociada a la condición de informalidad mediante la ecuación de Mincer y su validación a través de la descomposición de Oaxaca-Blinder, contrastando ambos enfoques para verificar la convergencia y robustez de sus resultados, de manera que permitan determinar si dicha penalización constituye una regularidad del mercado laboral ecuatoriano y no un artefacto de la técnica de estimación empleada, utilizando la ENEMDU 2025

Objetivo específico 3.- Identificar los perfiles socioeconómicos con mayor riesgo o probabilidad de caer en la informalidad laboral mediante regresión logística y Gradient Boosting Classifier, contrastando ambos enfoques con métricas de desempeño (precisión, recall, F1, AUC-ROC) y explicando sus determinantes a nivel individual mediante la técnica LIME.

Capítulo 2: Marco Teórico

Historia del estudio de la informalidad

Abordar el estudio de la economía informal partiendo de una correcta contextualización histórica, brinda mayores luces para dimensionar la real importancia de esta temática en el desarrollo de los estados. En este sentido, ¿por qué el mundo empezó a hablar de la informalidad? ¿Cuáles son sus antecedentes? ¿Cuáles fueron las causas que propiciaron el estudio del fenómeno? ¿Y qué desafíos plantea el estudio de la informalidad?

Responder a esas preguntas exige recorrer, aunque sea de forma sucinta, más de cinco décadas de producción académica e institucional sobre la informalidad laboral y su dinámica. Ese recorrido , que inicia desde la acuñación del término de "sector informal" por la Organización Internacional del Trabajo en su informe sobre Kenia en 1972, pasando por el hallazgo de Keith Hart (1973) que cuestionó la visión negativa inicial sobre su estudio, los debates de la década de 1980 entre la escuela neomarxista y la Escuela de Chicago, hasta la propuesta legalista de Hernando de Soto en 1989 y la posterior consolidación de la informalidad como campo de investigación interdisciplinario en América Latina; se desarrolla con mayor profundidad en el Anexo II de este trabajo, al que se remite al lector interesado en una lectura ampliada de cada uno de estos hitos y de las tesis de los distintos autores más influyentes que las sostienen.

Con el propósito de dotar al presente trabajo de una lectura ágil, ese recorrido histórico se condensa a continuación en dos líneas de tiempo (Figuras 1 y 2) que sintetizan visualmente los hitos, autores y escuelas de pensamiento reseñados en el Anexo II, organizados en dos tramos: 1949-1989, correspondiente a la génesis del concepto y su primera consolidación institucional; y 1990-2026, correspondiente a su expansión hacia enfoques legalistas, estructurales y, más recientemente, cuantitativos. Estas líneas de tiempo no son un simple recurso ilustrativo: constituyen el mapa conceptual que permite ubicar el enfoque legalista adoptado en este trabajo dentro de una tradición de pensamiento más amplia, y entender por qué las preguntas que hoy se formula la investigación sobre informalidad son, en gran medida, herederas de los debates y preguntas abiertas que ha dejado su estudio histórico y que se resumen a continuación:

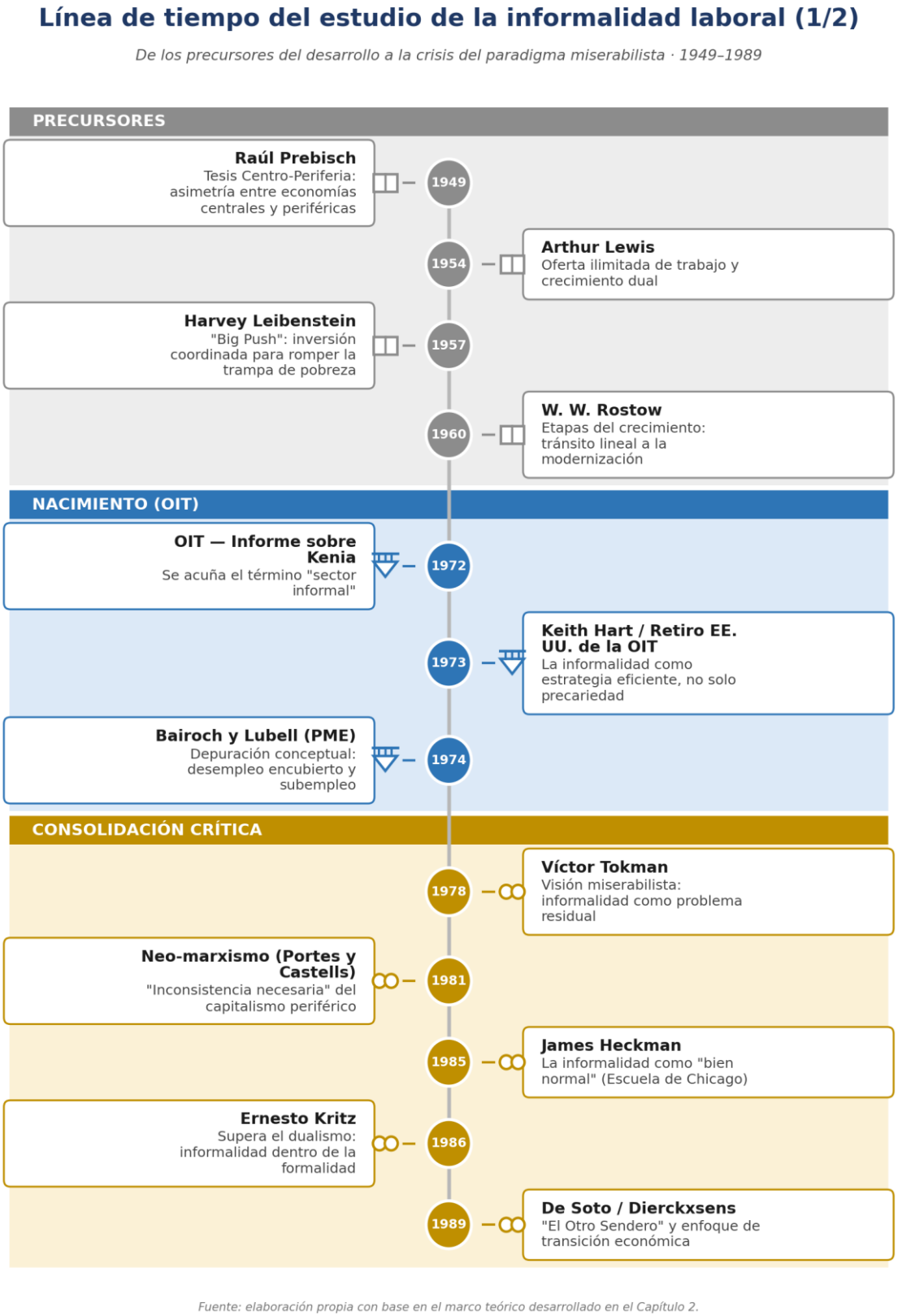


Figura 1. Línea de tiempo del estudio de la informalidad laboral (1949–1989). Fuente: elaboración propia con base en el recorrido histórico desarrollado en este capítulo.

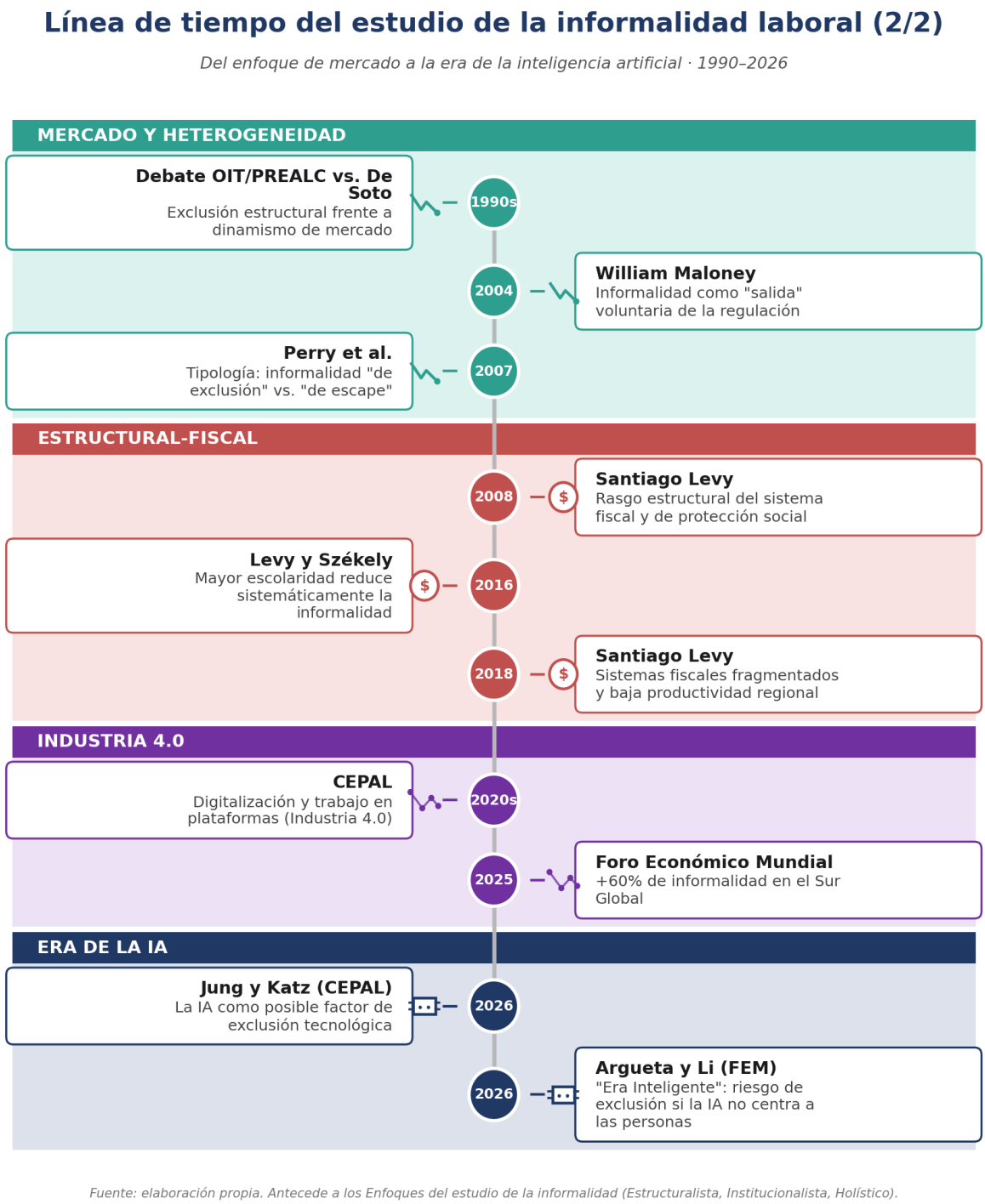


Figura 2. Línea de tiempo del estudio de la informalidad laboral (1990–2026). Fuente: elaboración propia con base en el recorrido histórico desarrollado en este capítulo.

La línea de tiempo presentada repasa los principales autores que, en distintos momentos, realizaron contribuciones relevantes al estudio de la informalidad en América Latina y en economías en vías de desarrollo. Sin embargo, este recorrido académico no refleja necesariamente una trayectoria lineal de avances acumulados, sino más bien una sucesión marcada por tensiones y contradicciones teóricas. La evolución del estudio de la informalidad laboral se presenta fragmentada, sin una continuidad evidente, lo que obliga a interpretar sus aportes como respuestas parciales a contextos específicos. Precisamente por ello, resulta necesario presentar los enfoques que han emergido de este recorrido histórico, pues constituyen fundamentos teóricos y síntesis conceptuales que permiten hoy en día ordenar y comprender la heterogeneidad del fenómeno.

En este sentido, el recorrido histórico presentado, más allá de su valor documental, cumple una función específica dentro de la arquitectura de este trabajo: explica por qué la informalidad admite lecturas tan distintas entre sí, como efecto residual y transitorio, como estrategia cotidiana y recurrente de subsistencia racional, o como resultado de una externalidad evitable ante el exceso de regulación estatal, y por qué ninguna de ellas ha logrado imponerse por completo sobre las demás. Cada enfoque que se revisa en la sección siguiente de este marco teórico (estructuralista, institucionalista, legalista, holístico) no surge de forma aislada, sino que responde, en mayor o menor medida, a alguno de los momentos reseñados en estas líneas de tiempo. Comprender esa genealogía es lo que permite justificar, más adelante, por qué este trabajo adopta específicamente el enfoque legalista y no una alternativa distinta, como marco de medición.

Enfoques del estudio de la informalidad:

Enfoque Estructuralista:

El enfoque estructuralista, desarrollado a partir de los aportes de Moser (1978), Castells y Portes (1989; 1993), Benton (1989) y Schauffler (1993), sostiene que la informalidad es un fenómeno de origen estructural, vinculado tanto a la dinámica de la oferta laboral como a la de la demanda laboral. No se trata de un problema coyuntural, sino de un desequilibrio persistente entre el crecimiento de la fuerza de trabajo y la limitada capacidad de absorción del sector formal, propio de las economías en vías de desarrollo.

Oferta laboral: sobre expansión y presión estructural

Los estructuralistas identifican tres factores principales que explican la sobreoferta de trabajo:

1. **Transición demográfica:** La reducción de la mortalidad infantil y el mantenimiento de altas tasas de natalidad incrementan la población económicamente activa. Sin embargo, gran parte de esta fuerza laboral carece de capacitación técnica, lo que limita su inserción en empleos formales.

2. **Migración rural-urbana:** Motivada por la caída de la demanda agrícola, el deterioro de salarios rurales y la percepción de mejores oportunidades en las ciudades. Aunque la migración se concibe como una mejora potencial del bienestar, en la práctica los migrantes enfrentan barreras de entrada al sector formal y recurren a la informalidad como estrategia de subsistencia (Longhi, 2007).

3. **Participación laboral familiar:** Ante ingresos insuficientes y restricciones presupuestarias, las familias incrementan el número de miembros activos o las horas trabajadas. Este ajuste redistributivo genera una presión adicional sobre el mercado laboral, que no logra absorber la nueva oferta (Perry & Maloney, 2007).

Estos factores muestran cómo la informalidad surge como respuesta adaptativa ante la incapacidad del sector formal para absorber la creciente fuerza laboral.

Demanda laboral: insuficiencia y rigidez estructural

Del lado de la demanda, la informalidad se explica por tres causas centrales:

- **Baja inversión y escasa innovación:** La insuficiencia de inversión en las economías en desarrollo constituye uno de los factores estructurales más relevantes para explicar la persistencia de la informalidad. La estrechez del mercado interno, derivada de bajos niveles de consumo y de ahorro, limita la capacidad de las empresas para expandirse y generar empleo formal. A ello se suma la fuga de capitales y la dificultad de inserción en los mercados internacionales, lo que restringe la disponibilidad de recursos para proyectos productivos de largo plazo. En este contexto, las empresas suelen optar por estrategias defensivas, privilegiando la reducción de costos antes que la creación de nuevas plazas de trabajo.

- El **atraso tecnológico** refuerza esta dinámica: la falta de innovación y de inversión en investigación y desarrollo impide que las industrias aumenten su productividad y diversifiquen sus procesos. Como consecuencia, la demanda laboral se mantiene rígida y de bajo crecimiento, incapaz de absorber la creciente oferta de mano de obra. Este círculo vicioso se traduce en una

mayor presión sobre la informalidad, que se convierte en el espacio donde la fuerza laboral excedente encuentra alternativas de subsistencia. En términos estructurales, la baja inversión y la escasa innovación reflejan una lógica de acumulación que prioriza la estabilidad financiera a corto plazo por encima de la expansión productiva a largo plazo. En ausencia de políticas públicas que incentiven la inversión, promuevan la innovación y fortalezcan la competitividad, la economía formal permanece limitada, y la informalidad se consolida como un mecanismo de ajuste frente a la incapacidad del sistema para generar empleo suficiente y de calidad (Longhi, 2007).

- **Composición del capital:** La estructura del capital en las economías en desarrollo suele privilegiar la inversión en capital fijo (maquinaria, infraestructura, tecnología) por encima de la inversión en capital variable (mano de obra). Este desbalance genera un proceso de desplazamiento de trabajadores**,** pues las empresas buscan aumentar la producción en serie y reducir los costos unitarios, pero lo hacen a expensas de la contratación formal. Como consecuencia, se produce un aumento del desempleo y una expansión de las actividades informales, que actúan como mecanismos de absorción de la mano de obra excedente. Este fenómeno se ve reforzado por prácticas como el maquilismo**,** la tercerización**,** la flexibilización laboral, o la digitalización no regulada de los bienes y servicios, que permiten a las empresas externalizar riesgos y costos, pero al mismo tiempo precarizan las condiciones de trabajo. En términos estructurales, la composición del capital refleja una lógica de acumulación que prioriza la eficiencia productiva sobre la estabilidad laboral, y que, en contextos de baja innovación y débil regulación estatal, contribuye directamente a la reproducción de la informalidad.

- **Ausencia de políticas públicas proactivas:** La falta de estrategias fiscales y laborales sostenidas constituye un factor decisivo en la perpetuación de la informalidad. En lugar de diseñar políticas de largo plazo orientadas a fortalecer la capacidad de absorción del mercado laboral, los gobiernos suelen aplicar medidas coyunturales que generan desplazamientos constantes entre la formalidad y la informalidad. Esta inestabilidad se traduce en una estructura laboral frágil, en la que los trabajadores carecen de seguridad laboral y las empresas operan sin incentivos claros para formalizar sus actividades. La ausencia de políticas proactivas también limita la capacidad del Estado para contrarrestar los efectos de las crisis económicas, pues el gasto público se utiliza de manera reactiva y no como herramienta de planificación estructural. En consecuencia, la informalidad se consolida como un mecanismo de supervivencia frente a la debilidad institucional, lo que refuerza la precariedad y amplía las brechas sociales en economías en vías de desarrollo.

En este marco, la informalidad no es un residuo del sistema, sino una consecuencia directa de la insuficiencia estructural de la demanda laboral. Por lo tanto, la población informal se caracteriza por tener bajos ingresos, escasa protección social, inestabilidad laboral, baja escolaridad y condiciones de trabajo deficientes (Portes, 2004). Estas características actúan como factores de autorrefuerzo: mientras los trabajadores formales mejoran su productividad y competitividad a lo largo de su carrera, los informales reproducen la precariedad y las limitaciones, lo que consolida brechas intergeneracionales (Uribe, 2004).

Enfoque Institucionalista:

Los máximos representantes de este enfoque son Michael Piore (1983), Peter Doeringer (1971) y Hernando de Soto (1989). Su aporte central consiste en explicar la informalidad a partir del papel de las instituciones estatales, entendidas como el conjunto de normas, reglas y parámetros de conducta que regulan la interacción social y económica. En este marco, la informalidad se interpreta como el resultado del incumplimiento parcial o total de dichas regulaciones, ya sean laborales, fiscales o municipales (Longhi, 2007).

El institucionalismo amplía el espectro de análisis al considerar no solo las leyes y decretos con respaldo estatal, sino también las convenciones internas de las empresas y los diseños organizativos que rigen el trabajo. Así, la estructura del mercado laboral se concibe como producto del patrón normativo dominante. La informalidad, entonces, no se reduce a una condición de marginalidad o de pobreza, sino que se entiende como una respuesta racional de los agentes económicos frente a la rigidez institucional y a los costos de la formalización.

A diferencia del estructuralismo, que apunta a factores demográficos o migratorios como causas de este fenómeno, el institucionalismo sitúa en el centro del debate la racionalidad costo-beneficio de los individuos. En este sentido, la informalidad se concibe como una decisión consciente y estratégica: los agentes evalúan los costos de cumplir la ley frente a los beneficios de evadirla y, en muchos casos, optan por la segunda alternativa. En palabras de De Soto (1989), la informalidad constituye una "alternativa bienvenida" para empresarios y trabajadores que buscan sobrevivir ante la excesiva regulación y los altos costos de la formalidad.

En este contexto, el enfoque institucionalista identifica tres factores principales como potenciales causas de la informalidad laboral:

1. **Barreras de entrada:** La concentración de mercados y la reglamentación excesiva constituyen obstáculos estructurales que limitan el acceso de nuevos actores al sector formal. En economías donde predominan poderes monopólicos u oligopólicos, los costos de participación, como licencias, trámites burocráticos, requisitos fiscales y laborales, se convierten en barreras que desalientan la formalización. Estas restricciones no solo elevan los costos de la inversión inicial, sino que también generan un entorno de exclusión en el que las pequeñas empresas y los trabajadores independientes encuentran difícil competir bajo las reglas del mercado formal.

   En este contexto, la informalidad emerge como una estrategia de democratización económica: permite que nuevos agentes ingresen a la actividad productiva sin enfrentar los elevados costos de entrada, diversificando la oferta y corrigiendo fallas de mercado asociadas a la concentración. Desde la perspectiva institucionalista, este fenómeno no debe interpretarse únicamente como precariedad, sino también como un mecanismo de apertura y competencia, que facilita la asignación más flexible de recursos y la desconcentración de mercados dominados por grandes actores.

   La paradoja radica en que las mismas instituciones diseñadas para garantizar el orden y la estabilidad (leyes, regulaciones y normativa secundaria) terminan por generar incentivos a la evasión. Cuanto más rígido y costoso es el acceso al mercado formal, mayor es la propensión de los agentes a recurrir a alternativas informales. Así, la informalidad se convierte en un espacio de innovación y supervivencia, pero también en un síntoma de la incapacidad institucional para equilibrar la eficiencia regulatoria con la inclusión económica.

2. **Costos de producción:** El cumplimiento de las obligaciones fiscales, laborales y municipales constituye uno de los factores más determinantes en la explicación institucionalista de la informalidad. En el sector formal, estas obligaciones se traducen en costos adicionales de producción que incluyen contribuciones a la seguridad social, impuestos municipales, regulaciones laborales, seguros y prestaciones obligatorias. Tales costos, aunque diseñados para garantizar derechos y estabilidad, elevan significativamente el precio final de los bienes y servicios, lo que reduce la competitividad de las empresas formales frente a sus pares informales.

   Las empresas informales, al prescindir de estos costos, logran estructuras de producción más flexibles y menos onerosas, lo que les permite ofrecer precios más bajos y adaptarse con rapidez a las fluctuaciones de la demanda. Este diferencial de costos genera un incentivo directo para que consumidores y familias opten por bienes y servicios informales, lo que refuerza la expansión del sector.

3. **Modelo tecnológico:** La búsqueda de eficiencia en las empresas formales ha conducido históricamente a procesos de tercerización y flexibilización laboral, configurando un modelo dualista de "centro y periferia". El centro, compuesto por trabajadores altamente calificados y protegidos, se beneficia de condiciones estables, del acceso a tecnología avanzada y de marcos institucionales sólidos. La periferia, en cambio, está conformada por trabajadores precarizados, tercerizados o externalizados, que sostienen la productividad del centro sin gozar de sus beneficios.

   En el contexto de la Industria 4.0 y de la inteligencia artificial, esta lógica se intensifica. Las empresas que adoptan tecnologías como la automatización, el big data, la robótica colaborativa y la IA tienden a concentrar las oportunidades en el núcleo de sus trabajadores especializados en programación, análisis de datos y gestión tecnológica. Mientras tanto, la periferia se amplía hacia formas de empleo más frágiles: trabajo en plataformas digitales, microtareas en línea, servicios externalizados mediante aplicaciones y actividades informales vinculadas a cadenas de valor digitales.

   La paradoja es que la superioridad tecnológica del centro depende de la periferia informal: la eficiencia y la productividad de los sectores altamente tecnificados se sostienen en una base laboral flexible, barata y desprotegida. La IA, al optimizar procesos y reducir costos, refuerza esta dualidad: por un lado, genera empleos de alta especialización; y por otro, desplaza a trabajadores poco calificados hacia la informalidad digital. Así, la informalidad ya no se limita a la venta callejera o al trabajo precario tradicional, sino que se expande hacia nuevas modalidades como el gig economy (economía de las plataformas), el trabajo en plataformas y la prestación de servicios digitales sin regulación.

   En términos agregados, el modelo tecnológico contemporáneo reproduce un patrón estable y selectivo para la élite empresarial, mientras resulta desarticulador e inestable para la periferia laboral. La informalidad se convierte en un componente funcional del sistema productivo digital: absorbe la mano de obra excedente, sostiene la competitividad del centro y refleja la incapacidad de los marcos institucionales para garantizar la inclusión en la era de la IA y la Industria 4.0.

   En consecuencia, desde la óptica institucionalista, este fenómeno no puede entenderse únicamente como resultado de la dinámica tecnológica, sino también como consecuencia de la acción o inacción de las instituciones estatales frente a dicho dinamismo. La ausencia de

regulaciones adaptadas a las nuevas formas de empleo digital, la rigidez de los marcos normativos tradicionales y la falta de políticas públicas proactivas generan un entorno en el que la informalidad digital prospera. En otras palabras, la expansión del trabajo en plataformas, la tercerización tecnológica y la precarización asociada a la gig economy no son inevitables, sino que reflejan la incapacidad del Estado para diseñar instituciones inclusivas que acompañen la transformación productiva.

Así, el institucionalismo ofrece una lectura crítica: la informalidad en la Industria 4.0 no surge únicamente de la lógica de la acumulación tecnológica, sino también de la brecha entre la velocidad de la innovación y la lentitud de la institucionalidad. Cuanto más rígidas y desactualizadas son las normas, mayor es el incentivo para que las empresas y los trabajadores operen en esquemas informales. El desafío contemporáneo, por tanto, consiste en construir instituciones capaces de equilibrar la eficiencia económica con la protección social, evitando que la revolución digital reproduzca y amplifique las desigualdades históricas del mercado laboral.

En consecuencia, el enfoque institucionalista aporta una visión distinta al debate: la informalidad no se concibe como un problema a erradicar, sino como un mecanismo funcional de los mercados, capaz de diversificar la competencia y de reducir las ineficiencias asociadas al costo de la ley. Sin embargo, esta interpretación revela una paradoja: cuanto más rígida es la institucionalidad estatal, mayor es el incentivo a evadirla, lo que refuerza la expansión del sector informal.

Enfoque Holístico:

El enfoque holístico parte del reconocimiento de que la informalidad es un fenómeno complejo y heterogéneo que no puede explicarse de manera suficiente con un único marco teórico. En lugar de concebirla como homogénea, este enfoque reconoce la coexistencia de múltiples modalidades y matices, lo que exige integrar las perspectivas estructuralistas e institucionalistas en un esfuerzo armonizado.

Mientras el estructuralismo subraya los factores demográficos, migratorios y económicos que expulsan a los individuos hacia la informalidad, el institucionalismo enfatiza la racionalidad costo-beneficio y la rigidez normativa que incentivan la evasión de la formalidad. El enfoque holístico sostiene que ambos marcos no son excluyentes, sino complementarios, y que solo su síntesis permite comprender la informalidad como un fenómeno integral. En este contexto, la informalidad se interpreta a partir de una disyuntiva entre exclusión y escape:

- **Exclusión:** La estructura económica, social y cultural expulsa a individuos vulnerables con baja escolaridad, escasa tecnificación y bajos ingresos hacia la informalidad, privándolos de acceso a la economía formal.

- **Escape:** Otros agentes, en cambio, optan por la informalidad de manera consciente y estratégica, como alternativa para evadir los costos de la formalización y la presión regulatoria del Estado.

El tamaño y la eficiencia de la institucionalidad estatal inciden directamente en esta dinámica. Cuanto más rígidas y costosas son las regulaciones, mayor es el incentivo para el “escape”; cuanto más débiles y fragmentadas son las políticas públicas, mayor es la probabilidad de “exclusión”. Así, la informalidad se configura como un espacio donde convergen tanto la falta de oportunidades estructurales como la decisión racional de evadir la formalidad.

Delimitación de informalidad:

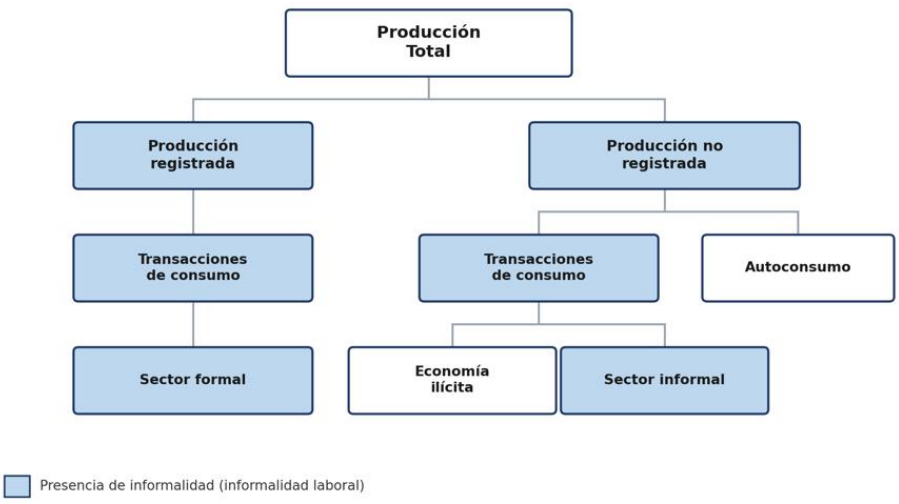
La revisión histórica del estudio de la informalidad resulta fundamental para comprender sus antecedentes, la relevancia de su análisis y la diversidad de perspectivas desde las que se ha abordado. No obstante, para enfrentar con claridad esta problemática es indispensable contar con una definición precisa del fenómeno, delimitación que ningún enfoque logra cubrir de manera exhaustiva. Con el fin de reducir ambigüedades, la literatura introduce una distinción clave: la informalidad constituye un universo amplio, dentro del cual el empleo informal o informalidad laboral representa solo uno de sus componentes, presente de manera transversal en distintos sectores de la economía, tanto formales como informales. A partir de esta precisión, se han desarrollado dos aproximaciones que permiten delimitar su estudio:

- **Definición legalista:** identifica como informales a las empresas que no cumplen con la normativa laboral o fiscal y cuyos trabajadores carecen de derechos de protección social. En esta línea, la OIT caracteriza al trabajador informal como aquel cuya relación de empleo no está regulada por la legislación laboral que, además, no accede a mecanismos de seguridad social (Gasparini & Tornarolli, 2010) o que labora en empresas que no cuentan con registros contables.

- **Definición productivista:** asocia la informalidad con unidades económicas de baja productividad, escaso capital y uso limitado de tecnología. Bajo este enfoque, integran la informalidad laboral establecimientos pequeños, con poca absorción de fuerza laboral y que desempeñan actividades con alta participación de mano de obra poco calificada, y que emplean generalmente técnicas rudimentarias en sus procesos productivos, careciendo de innovación tecnológica (Maloney, 2004).

En este marco, la informalidad que se abordará en la presente investigación, no sólo tiene lugar en el sector informal[1], sino que se extiende al sector formal e implica profundas interrelaciones entre ambos (ver tabla número 3):

Figura 3. Extensión y delimitación de la informalidad.



Fuente: Oscar Enríquez (2014)
Elaboración: Propia

[1] Sector informal: conjunto de actividades productivas que tienen lugar en la economía no registrada y que no son ilegales.

Tal como se observa en la tabla anterior, la producción puede dividirse en dos grandes categorías: registrada y no registrada. La primera corresponde al sector formal, integrado por actividades que cumplen con la legislación y la normativa vigentes. La segunda incluye tanto actividades que implican transacciones de consumo, es decir, intercambios comerciales, como el autoconsumo, entendido como el conjunto de actividades productivas destinadas a satisfacer directamente las necesidades de quienes las generan, sin mediar intercambio en el mercado.

Dentro de las transacciones de consumo se distinguen dos sectores: el ilícito, que agrupa actividades cuya producción y comercialización están prohibidas por la normativa, y el informal, conformado por actividades legales en esencia, pero que se llevan a cabo mediante prácticas que vulneran la legislación (por ejemplo, evasión fiscal, ausencia de contratos laborales, trabajo no remunerado, etc.).

Es importante subrayar que la informalidad no se limita exclusivamente al sector informal. En realidad, puede manifestarse en distintos ámbitos de la producción, incluso dentro de la formalidad, dependiendo de la definición que se adopte, ya sea la productivista o la legalista. Esto evidencia la existencia de interrelaciones y sinergias entre ambos sectores, lo que refuerza la idea de que la informalidad es un fenómeno transversal, especialmente con el auge de la gig economy y de la economía digital.

Tipología funcional:

La informalidad es una problemática compleja y, en términos del grupo poblacional, este es heterogéneo y podría emplearse una multiplicidad de criterios para tipificar sus características. Sin embargo, continuando con el esfuerzo de Enríquez (2014), y tomando como punto de partida el grado de autonomía y disponibilidad del capital humano, se obtiene una tipología útil y sencilla, que permite segmentar al grupo informal para abordar de mejor manera su estudio. En este sentido, se presentan los siguientes tipos de informalidad:

- **Informalidad cuentapropista:** En este grupo se sitúan todos aquellos agentes que cuentan con un establecimiento informal, independientemente de su tamaño. Esta categoría incluye emprendedores, dueños y socios. Se identifica a un establecimiento informal como aquel que no presenta registros contables, libros administrativos ni RUC, o cuyo tamaño es inferior a 10 empleados. (Enríquez, 2014).

- **Informalidad asalariada:** Incluye a todos los trabajadores que laboran en una relación de dependencia informal. Convencionalmente, una relación de dependencia informal se caracteriza por: ausencia de afiliación a la seguridad social, incumplimiento de obligaciones patronales e inexistencia de estabilidad laboral. (Enríquez, 2014).

- **Informalidad no remunerada:** Esta categoría de informalidad incluye a trabajadores que no reciben remuneración económica por su trabajo y se ven motivados a desempeñar su labor por una relación familiar o por relaciones solidarias. (Enríquez, 2014).

Capítulo 3: Metodología

Técnicas de Investigación aplicadas:

La presente investigación adopta un enfoque cuantitativo-inferencial, fundamentado en la recolección y el análisis de datos numéricos de encuestas de hogar, cuya unidad de observación es el hogar ecuatoriano en el año 2025 y cuya unidad de análisis es la persona de la Población Económicamente Activa (PEA) mayor de 18 años. Su enfoque es cuantitativo porque se enfoca en la recolección y análisis de microdatos recolectados a partir de encuestas de hogar estandarizados y armonizados en el tiempo con el fin de establecer relaciones entre condiciones socioeconómicas y efectos laborales derivados a partir de la precarización laboral sobre el mercado laboral. Es también una investigación inferencial, pues, haciendo uso de la construcción metodológica, del diseño muestral y del nivel de representatividad de la ENEMDU, busca extrapolar y generalizar los patrones de penalización salarial y los perfiles de riesgo encontrados en la muestra a toda la Población Económicamente Activa (PEA) de Ecuador, mediante sus factores de expansión.

Este enfoque es el que mejor se adapta a los objetivos del estudio, pues garantiza el rigor, la precisión, así como la validez interna y externa de la investigación, y la hace comparable y replicable con los esfuerzos regionales realizados en otros países de Sudamérica. Además, permite transitar desde la inferencia estadística lineal tradicional hacia un modelamiento computacional avanzado de Machine Learning, mediante el cual los resultados obtenidos en el país sobre perfiles de riesgo y el costo de oportunidad asociados a la informalidad, obtendrán un relevo al ser abordados a la luz de métodos sofisticados que aseguran la precisión de los resultados y que al momento no se han aplicado al estudio de este fenómeno. Con ello, las recomendaciones de política pública que se generen serán robustas, sólidas, relevantes y objetivas.

En este sentido, con base en las rutas metodológicas formales del ámbito científico, este enfoque se caracteriza por ser estructurado, replicable y apoyado firmemente en métodos estadísticos avanzados y técnicas inferenciales estandarizadas aplicados al fenómeno socioeconómico antes citado desde la ciencia de datos. Como se dijo, la suma de todos estos factores permitirá la generalización y la validez interna y externa de los resultados a partir de una muestra representativa, previamente validada, y de producción periódica por parte del Instituto Nacional de Estadística y Censos para su Operación Estadística “Encuesta Nacional de Empleo y Desempleo Urbano del Ecuador” (ENEMDU-2025).

Diseño de Investigación:

La arquitectura metodológica de este estudio se estructura bajo un diseño de investigación no experimental, de corte transversal, con un alcance correlacional-causal y predictivo:

- Naturaleza No Experimental: El estudio se define como no experimental, ya que las variables causales y derivadas de la informalidad, así como las variables empleadas para su definición y delimitación, constituyen características sociodemográficas de los individuos observadas en 2025, sin que hayan sufrido modificación alguna para responder a uno o varios objetivos de la

presente investigación. En este sentido, constituyen una fotografía fiel de la calidad y el estado situacional del sistema laboral ecuatoriano. En este contexto, el fenómeno de la informalidad y el entramado de relaciones socioeconómicas detrás de esta problemática serán observadas en su entorno real y de manera objetiva, tal como fueron recolectados por la operación estadística oficial, lo que permitirá un análisis posterior de sus interrelaciones sin alterar la naturaleza de los datos.

- Temporalidad Transversal: La investigación es transversal dado que el análisis se circunscribe a un momento único en el tiempo, utilizando como corte temporal los microdatos consolidados de la ENEMDU correspondientes a diciembre de 2025. Este enfoque temporal es el estándar idóneo para estimar la penalización salarial y los perfiles de riesgo en un escenario macroeconómico y laboral específico, proporcionando una "fotografía" analítica de alta fidelidad de la realidad estructural del país.

- Alcance Correlacional-Causal y Predictivo: El alcance supera la descripción estadística convencional para adentrarse en la evaluación de la causalidad y de la capacidad predictiva mediante Machine Learning, con el fin de determinar perfiles socioeconómicos con riesgo de caer en la informalidad. Es correlacional-causal porque busca determinar el impacto de la condición de informalidad en el salario de los trabajadores con experiencia y formación equivalentes, en comparación con individuos con experiencia y formación equivalentes que desempeñan su actividad en la economía formal. Este efecto se podrá observar, entre otras, en la variable de ingresos del trabajador (aislando el efecto del capital humano). Es predictivo porque, a partir del entrenamiento de algoritmos supervisados, se modela la probabilidad condicional de un individuo de pertenecer al sector informal en función de sus atributos individuales, estimando con precisión sus factores de vulnerabilidad.

Este diseño metodológico responde directamente a los tres objetivos específicos de la disertación. En primer lugar, la naturaleza predictiva y el alcance causal son indispensables para identificar los determinantes económicos y sociodemográficos que inciden en la probabilidad de inserción en el empleo informal. En segundo lugar, el carácter correlacional transversal permite comparar el desempeño predictivo y explicativo de la ecuación econométrica tradicional de Mincer y de la ecuación de Oaxaca-Blinder. Finalmente, este diseño permite analizar los matices de la penalización salarial según los perfiles de riesgo, facilitando la segmentación ex post de los resultados para identificar impactos heterogéneos por género, área geográfica, nivel educativo o por distintas condiciones socioeconómicas.

Técnica de recolección de información, muestreo e identificación de la población objetivo

Técnica de Recolección de Información:

La investigación se fundamenta en la recopilación de datos secundarios homologados. No se realizará levantamiento de campo directo; en su lugar, se accederá a microdatos a nivel de persona en formatos abiertos procesables (.csv, .sav y .dta) provenientes del repositorio oficial del Instituto Nacional de Estadística y Censos (INEC). La recolección de información original en campo efectuada por el INEC se rige por normas internacionales de captura de información, con las que se protege la seguridad y la privacidad de los datos personales del informante, se asegura una carga balanceada de respuestas, se asegura contactar con un informante calificado y se precauta contra la invasión del informante. Estos principios y técnicas, que garantizan la fiabilidad y la protección del microdato, figuran en el Manual del Encuestador de la ENEMDU. Lo descrito constituye una ventaja relevante de la investigación, pues se podrá acceder a información precisa de microdatos con alcance nacional, sin incurrir en tiempos muertos en el levantamiento de información y sin arriesgar la privacidad de los datos personales de los informantes, dando cumplimiento a lo dispuesto en la Ley de Datos Personales del país.

Así mismo, al ser la ENEMDU una encuesta de generación continua y periódica y contar con una certificación de Operación Estadística, ello garantiza un mínimo de calidad en el microdato, de modo que las preguntas y objetivos de la investigación puedan ser respondidos, sin riesgo de no contar con información o de que exista desactualización de datos. En el ámbito de la ciencia de datos, estas son dos ventajas comparativas relevantes; por ello, la investigación se plantea a la luz de las características de esta encuesta.

Muestra y población objetivo:

Universo de investigación: El universo de estudio de la ENEMDU son personas de 15 años o más de edad, residentes en las viviendas del Ecuador, exceptuando a la población que reside en viviendas colectivas, viviendas flotantes y población indigente (sin techo), que pertenezcan a la Población Económicamente Activa.

Unidad de observación: La unidad de observación comprende todas las viviendas particulares ocupadas que se encuentran en territorio nacional, las cuales tienen ligada su identificación geográfica mediante fuentes cartográficas.

Unidad de análisis: Para los indicadores laborales, en especial para la informalidad, la población de referencia es la de todas las personas mayores de 18 años.

Cobertura geográfica: La cobertura geográfica está definida por las viviendas ocupadas ubicadas en el territorio ecuatoriano, incluida la región insular.

Tipo de muestreo: El de la ENEMDU para el estudio de la informalidad es un muestreo probabilístico estratificado bietápico de elementos.

Dominios de estudio y representatividad: La ENEMDU anual tiene como dominios de estudio y representatividad nacional: Urbano-Rural, 5 ciudades principales (Quito, Guayaquil, Cuenca, Machala y Ambato) y 24 provincias del Ecuador.

Tamaño y selección de la muestra para el Estudio de la Informalidad: La confiabilidad estadística de la encuesta base (ENEMDU) es del 95%. La expresión matemática utilizada para el cálculo de tamaño de muestra de personas en cada dominio de diseño D es:

$nD \geq (P(1-P)\ Deff)/\ \{(\delta\ 2P2\ /z\alpha\ 2\ )+\ (P(1-P)\ Deff/\ ND)\ \}*\ (1/\ (1-TN\mathrm{R})$

Donde:

z = $percentil\ de\ la\ distribución\ normal\ estándar\ asociado\ al\ nivel\ de\ confianza\ 1-\alpha$

$\delta = margen\ de\ error\ relativo\ máximo$

$P = estimación\ de\ la\ variable\ de\ diseño\ (tasa\ de\ desempleo)$

$ND = tamaño\ de\ la\ población\ en\ cada\ dominio\ de\ diseño\ D$

$TNR = tasa\ de\ no\ respuesta$

Cuadro 1. Diseño muestral. Informalidad en la ENEMDU 2025

Dominio	**Viviendas del Marco**	**Conglomerados del Marco**	**Conglomerados incluidos en la muestra**	**Viviendas incluidas en la muestra**
Urbano	2.715.812	79.647	11.424	79.968
Rural	1.002.506	27.352	4.032	28.224

Fuente: INEC, ENEMDU,

Elaboración: Óscar Enríquez.

Con base en este diseño muestral, la población analizada será de 8,6 millones de personas pertenecientes a la Población Económicamente Activa en Ecuador. De esta población se analizará específicamente a 4’275.784 personas que presentan condición de informalidad a diciembre de 2025.

5. Principales variables de análisis y técnicas de recolección:

Como se indicó en la sección anterior, la investigación se fundamenta en la recopilación de datos secundarios homologados. No se realizará levantamiento de campo directo; en su lugar, se emplearán como fuentes de información los archivos institucionales de microdatos en formatos abiertos procesables (.csv, .sav o .dta) del repositorio oficial del INEC. La recolección original en campo, efectuada por el INEC, se rige por el Manual del Encuestador de la ENEMDU y utiliza entrevistas directas en el hogar mediante dispositivos móviles de captura de datos (CAPI).

A partir de la estructura del formulario oficial de la ENEMDU, se seleccionarán y construirán los siguientes vectores de variables:

Tipo de Variable	Nombre de la Variable	Indicador / Definición Operacional	Relación con ENEMDU (Preguntas/Secciones)
Variable Dependiente (Regresión)	Log_Ingreso	Logaritmo natural del ingreso laboral por hora de la ocupación principal.	Sección de Ingresos (Ocupación principal y horas trabajadas).
Variable Dependiente (Clasificación)	Condicion_Informal	Variable binaria (\(1 = \text{Informal}\), \(0 = \text{Formal}\)). Definida bajo el criterio legalista (no afiliación a la seguridad social del IESS) (Seminario ... p. 6, El reto de... p. 28).	Sección de Ocupación (Pregunta sobre la aportación al seguro social).
Variables Estructurales (Capital Humano)	Educacion / Experiencia	Años de escolaridad formal aprobada y experiencia laboral potencial (\(\text{Edad} - \text{Educación} - 6\)) (Seminario ... p. 6).	Sección de Educación y Datos Demográficos Generales.
Variables Estructurales (Demográficas)	Genero / Edad / Etnia	Vectores sociodemográficos de exclusión latente (Variables de control y segmentación) (Seminario ... pp. 2, 7).	Sección de Características de los Miembros del Hogar.
Variables Coyunturales (Laborales)	Tamaño_Firma	Número de trabajadores en el establecimiento o unidad económica (\(<10\) empleados).	Sección de Características de la Ocupación Principal.
Variables Coyunturales (Laborales)	Tipo_Contrato	Condición contractual del trabajador (asalariado con/sin contrato, jornalero, cuenta propia).	Sección de Características de la Ocupación Principal.
Variables Institucionales / Fiscales	Rama_Actividad	Clasificación industrial del establecimiento (CIIU) a nivel de dígitos permitidos.	Sección de Características de la Ocupación Principal.
Variables Institucionales / Fiscales	Regimen_Impositivo	Condición de registro fiscal del negocio (declaración de impuestos o tenencia de RUC/RIMPE).	Sección de Características de la Ocupación Principal.
Variables de Control (Demográficas)	Genero / Area_Geografica	Condiciones demográficas asociadas al individuo.	Sección de Características de los Miembros del Hogar.

Fuente: INEC (ENEMDU)

Elaboración: Óscar Enríquez.

Modelo analítico y métricas

El procesamiento de datos y modelamiento se ejecutará en lenguajes de programación científica (Python o R), dividiendo el flujo analítico en tres fases secuenciales integradas:

Fase 1: Estimación de la Penalización Salarial Asociada a la Informalidad

Para aislar la brecha salarial neta atribuible a la condición de informalidad, más allá de la diferencia bruta que arroja la estadística descriptiva, se emplearán dos enfoques complementarios: la ecuación de Mincer extendida y la descomposición de Oaxaca-Blinder. Ambos parten de la misma teoría del capital humano, pero difieren en el supuesto que asumen sobre la estructura de retornos del mercado laboral, lo que permite someter el hallazgo de uno a la verificación del otro.

Modelo de Mincer Extendido

La ecuación de Mincer explica el logaritmo del ingreso laboral de un trabajador a partir de su capital humano y de un conjunto de características estructurales adicionales que la especificación clásica, centrada únicamente en la escolaridad y la experiencia, no captura. La forma extendida que se empleará en este trabajo adopta la siguiente estructura general:

$$ln(ingreso) = \beta_0 + \beta_1 \cdot informal + \beta_2 \cdot escolaridad + \beta_3 \cdot experiencia + \beta_4 \cdot experiencia^2 + \beta_5 \cdot sexo + \beta_6 \cdot área + \Sigma\gamma_k \cdot estado_civil_k + \Sigma\delta_j \cdot tamaño_empresa_j + \Sigma\theta_m \cdot ocupación_m + \Sigma\rho_r \cdot rama_r + \Sigma\varphi_\varphi \cdot quintil_\varphi + \varepsilon$$

donde cada término representa lo siguiente: informal es una variable dicotómica que distingue a los trabajadores informales de los formales, y cuyo coeficiente constituye el principal objeto de estimación de esta fase; escolaridad recoge los años de educación formal aprobados; experiencia y su término cuadrático experiencia² aproximan la trayectoria de ingresos a lo largo de la vida laboral, incluyendo el rendimiento decreciente que la teoría del capital humano predice; sexo y área son variables dicotómicas de control demográfico y territorial; estado_civil, tamaño_empresa, ocupación, rama y quintil son conjuntos de variables categóricas, cada una definida frente a una categoría de referencia, que controlan por la estructura del hogar, la escala del establecimiento, la inserción ocupacional, el sector productivo y la posición socioeconómica del hogar de origen, respectivamente. El vector β agrupa los coeficientes de las variables continuas y dicotómicas simples, mientras que γ, δ, θ, ρ y φ agrupan los coeficientes de cada categoría dentro de sus respectivos grupos de variables categóricas.

La interpretación de cada coeficiente sigue una regla común a toda la ecuación: mide el efecto marginal de la variable correspondiente sobre el logaritmo del ingreso, manteniendo constante el resto de covariables del modelo. Para las variables categóricas, el coeficiente siempre se lee en comparación con su categoría de referencia -aquella que no recibe una variable dicotómica propia y que queda representada en el intercepto-; un coeficiente positivo indica que esa categoría se asocia con un ingreso mayor que la categoría de referencia, y uno negativo, con un ingreso menor, en ambos casos manteniendo el resto de características constantes. Dado que la variable dependiente está expresada en logaritmos, ningún coeficiente se interpreta directamente como un monto en dólares ni como un porcentaje; para ello se requieren las transformaciones descritas a continuación.

El logaritmo del ingreso puede convertirse de vuelta a su magnitud monetaria mediante la función exponencial inversa a la transformación empleada en la construcción de la variable dependiente:

$$ingreso\ (USD) = exp(ln(ingreso)) - 1$$

y el efecto porcentual de cualquier coeficiente β sobre el ingreso -es decir, la brecha expresada como una tasa de variación, y no como una diferencia en logaritmos- se obtiene exponenciando el coeficiente estimado:

$$\%\Delta\ ingreso = (e^{\beta} - 1) \times 100$$

Esta última transformación será la que se utilice, en el capítulo de resultados, para traducir el coeficiente de la variable informal -y el de cualquier otra covariable de interés- a un lenguaje directamente interpretable como penalización o prima salarial porcentual.

Descomposición de Oaxaca-Blinder

La ecuación de Mincer extendida descansa sobre un supuesto que conviene poner a prueba: que los formales e informales comparten la misma estructura de retornos a sus características, y que la única diferencia entre ambos grupos es el intercepto, que captura la condición de informalidad. La descomposición de Oaxaca-Blinder permite relajar ese supuesto, al estimar una ecuación de ingresos independiente para cada grupo —formales e informales—, con la misma estructura de covariables de la ecuación extendida, pero sin la variable informal, que en este caso deja de ser necesaria porque define directamente a qué submuestra pertenece cada observación:

$$ln(ingreso)_F = X_F'\beta_F + \varepsilon_F \qquad ln(ingreso)_I = X_I'\beta_I + \varepsilon_I$$

donde los subíndices F e I identifican, respectivamente, a la ecuación estimada sobre la submuestra formal y sobre la submuestra informal, X representa el vector de covariables de cada grupo y β su correspondiente estructura de retornos. A partir de ambas ecuaciones, la brecha promedio de ingresos entre los dos grupos se descompone en dos componentes:

$$\Delta = (\bar{X}_F - \bar{X}_I)'\beta_F + \bar{X}_I'(\beta_F - \beta_I)$$

$$\Delta = Efecto\ dotaciones + Efecto\ coeficientes$$

El efecto dotaciones, primer término de la identidad, mide la parte de la brecha que se explica por el hecho de que los dos grupos difieren, en promedio, en sus características observables (escolaridad, experiencia, tamaño de establecimiento, entre otras), evaluadas con la estructura de retornos del sector formal. El efecto de coeficientes, segundo término, mide la parte de la brecha que persistiría incluso si ambos grupos tuvieran exactamente las mismas características, y que, por tanto, responde exclusivamente a que el mercado retribuye esas características de manera distinta según la condición de formalidad del trabajador; esta segunda cifra constituye, en sentido estricto, la penalización pura de la informalidad, libre del supuesto de retornos comunes que impone la ecuación extendida. Ambos componentes se expresan en la misma escala logarítmica que la variable dependiente, por lo que su interpretación como participación de la brecha total -dividiendo cada componente entre la brecha total— o su conversión a un efecto porcentual independiente -aplicando la transformación exponencial ya descrita— seguirá los mismos principios detallados para la ecuación de Mincer.

Fase 2: Clasificación de Perfiles de Riesgo de Informalidad

Para identificar qué características se asocian con una mayor probabilidad de que un trabajador se encuentre en condición de informalidad, se empleará un ejercicio de clasificación binaria que combina dos técnicas de naturaleza distinta: una regresión logística, que ofrece coeficientes directamente

interpretables, y un Gradient Boosting Classifier, que sacrifica esa interpretabilidad directa a cambio de una mayor flexibilidad para capturar relaciones no lineales e interacciones entre variables. El objetivo no es seleccionar un único modelo ganador, sino verificar si ambos enfoques convergen en señalar a los mismos determinantes como los más relevantes, lo que constituiría evidencia de que dichos determinantes son estructurales y no un artefacto de la técnica de estimación empleada.

Regresión Logística

La regresión logística modela el logaritmo de la razón de momios de informalidad como una función lineal de las mismas covariables estructurales empleadas en la Fase 1 —escolaridad, experiencia, sexo, área, estado civil, tamaño de establecimiento, ocupación, rama de actividad y quintil de ingreso del hogar—, de acuerdo con la siguiente forma funcional:

$$ln[p / (1 - p)] = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k$$

donde p es la probabilidad de que el trabajador sea informal y X_1 a X_k son las covariables del modelo, con la misma lógica de variables categóricas y categoría de referencia descrita en la ecuación de Mincer. La interpretación de los coeficientes, sin embargo, no se hace directamente sobre β, sino sobre su exponencial, conocida como razón de momios (odds ratio):

$$OR = e^{\beta}$$

Un odds ratio superior a uno indica que la categoría en cuestión incrementa la probabilidad relativa de informalidad frente a su categoría de referencia; un valor inferior a uno indica un efecto protector; y un valor igual a uno indica ausencia de asociación. Esta métrica será la que se reporte de forma sistemática en el capítulo de resultados de cada variable del modelo.

Gradient Boosting Classifier

El Gradient Boosting Classifier construye, de forma secuencial, un ensamble de árboles de decisión en el que cada árbol corrige los errores del ensamble construido hasta el paso anterior, de acuerdo con la forma general:

$$F(x) = \Sigma_{m=1}^{M} \gamma_m h_m(x)$$

donde cada $h_m(x)$ es un árbol de decisión simple ajustado a los residuos del modelo previo, γ_m es el peso con el que ese árbol se incorpora al resultado final, y M es el número total de árboles secuenciales, seleccionado mediante una comparación de hiperparámetros en el conjunto de prueba. A diferencia de la regresión logística, este modelo no impone una forma funcional lineal sobre el log-odds ni produce coeficientes interpretables por variable; en su lugar, la contribución de cada covariable se evalúa mediante medidas de importancia —típicamente, la reducción de impureza que cada variable aporta al conjunto de árboles—, que permiten establecer un ranking de relevancia comparable, aunque no directamente equivalente en escala, al de los coeficientes de la regresión logística. Ambos modelos se entrenarán sobre la misma partición de la muestra (70% entrenamiento, 30% prueba) y se evaluarán con las mismas métricas de desempeño —área bajo la curva ROC, precisión, recall y F1-score—, lo que garantiza que la comparación de sus resultados en el capítulo siguiente sea directa y metodológicamente consistente. Adicionalmente, la selección de hiperparámetros del Gradient Boosting Classifier —en particular, la profundidad máxima de los árboles— se apoyará en validación cruzada de k-folds (k=5) sobre el conjunto de entrenamiento, como resguardo adicional frente al sobreajuste antes de evaluar el desempeño final sobre el conjunto de prueba.

Fase 3: Análisis de la Heterogeneidad según Perfil de Riesgo Asociado a la Penalización Salarial

Los coeficientes y las medidas de importancia obtenidos en la Fase 2 describen, en el mejor de los casos, un efecto promedio de cada variable sobre el conjunto de la muestra. Esa lectura agregada, sin embargo, no garantiza que el mecanismo identificado se cumpla de la misma manera para todos los perfiles individuales, ni permite anticipar si existen combinaciones particulares de características —por ejemplo, un origen socioeconómico favorable combinado con una inserción ocupacional precaria— en las que el riesgo de informalidad se aparte de lo que el promedio sugeriría. Esta limitación resulta particularmente relevante para el modelo Gradient Boosting Classifier, que, por su propia naturaleza no lineal y basada en árboles de decisión, no ofrece un coeficiente único e interpretable por variable, sino que opera como una caja negra: produce predicciones de alta capacidad discriminante sin que sea posible, a simple vista, atribuir esa capacidad a razones concretas y verificables caso por caso.

Para resolver esta limitación se empleará la técnica LIME (Local Interpretable Model-agnostic Explanations), propuesta por Ribeiro, Singh y Guestrin (2016), sobre el modelo con mejor desempeño predictivo seleccionado en la Fase 2. LIME aproxima, para una observación puntual, el comportamiento del modelo complejo mediante un modelo lineal simple ajustado localmente alrededor de esa observación: perturba ligeramente los valores de las variables del caso de interés, observa cómo cambia la probabilidad predicha por el modelo original ante esas perturbaciones, y ajusta con esos resultados una regresión lineal ponderada que aproxima el comportamiento del modelo complejo únicamente en el entorno inmediato de esa observación —no en el conjunto de la muestra—. El resultado es una explicación tan interpretable como la de la regresión logística, pero construida individuo por individuo, lo que permite responder no solo cuáles son los determinantes agregados de la informalidad, sino por qué el modelo predice un riesgo determinado para un trabajador específico, y cuál de sus características -género, área geográfica, nivel educativo, quintil de ingreso del hogar, categoría ocupacional, entre otras- contribuye en mayor medida a esa predicción particular.

En la aplicación de esta fase se seleccionará un conjunto reducido de casos representativos del conjunto de prueba, elegidos deliberadamente para contrastar combinaciones de interés -por ejemplo, niveles extremos de escolaridad o de quintil de ingreso del hogar-, con el propósito de examinar si los determinantes que la explicación local atribuye a cada caso son consistentes con los hallazgos agregados de la Fase 2, o si revelan matices -como la posible dominancia de la categoría ocupacional sobre el origen socioeconómico en la explicación de casos específicos- que la sola lectura de coeficientes o de importancia de variables no permitiría anticipar. Este ejercicio de explicabilidad local constituye, junto con las dos fases anteriores, el insumo directo para identificar los perfiles de riesgo y las recomendaciones de política pública que se desarrollan en el capítulo de resultados.

Capítulo 4: Resultados

4.1. Dimensionamiento y Caracterización de la Informalidad Laboral en Ecuador

Antes de determinar si la informalidad laboral supone un costo cierto sobre el ingreso de los trabajadores y, en ese caso, identificar las condiciones que impulsan su proliferación en la vida productiva de la población económicamente activa, resulta indispensable dimensionar y caracterizar el fenómeno para comprender sus dimensiones reales y su multiplicidad de formas. Comprender cuántos ecuatorianos se encuentran en esta condición, dónde se concentran, en qué actividades productivas se insertan y qué atributos sociodemográficos comparten no constituye un esfuerzo meramente descriptivo: es el punto de partida que permite mapear, visibilizar e interpretar las dinámicas de un segmento poblacional que permanece invisible u oculto para la política pública. Es este punto de partida precisamente, el que permitirá responder al primer objetivo del presente trabajo de titulación y el factor que más adelante permitirá darle contexto a una eventual penalización salarial y constituirá el antecedente que dará sustento a la matriz de perfiles de riesgo de la informalidad, mediante la que se demostrará que la informalidad es un todo heterogéneo de amplio alcance en la PEA. Como nota conceptual, cabe destacar que los resultados que se presentan a continuación provienen del procesamiento de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) de 2025, publicada por el Instituto Nacional de Estadística y Censos (INEC). La información ha sido expandida mediante el factor de ponderación muestral (fexp), lo que permite obtener inferencias representativas a nivel nacional. Este procedimiento asegura que cada observación de la muestra refleje adecuadamente el peso poblacional correspondiente, corrigiendo posibles sesgos de selección y garantizando la consistencia con las proyecciones demográficas oficiales. Asimismo, se incorpora el balance muestral de las ciudades autorrepresentadas (Quito, Guayaquil y Cuenca), que, por su tamaño y relevancia estadística, se incluyen de manera diferenciada en el diseño de la encuesta. De este modo, los resultados no solo reflejan la estructura laboral nacional, sino también las particularidades urbanas de los principales centros económicos del país.

4.1.1. La magnitud del fenómeno: casi uno de cada dos trabajadores

En este contexto, el primer hallazgo es de escala. De una Población Económicamente Activa (PEA) estimada en 8'612.073 personas, un total de 4'275.784 se encuentran en situación de informalidad laboral, lo que arroja una tasa de informalidad agregada del 49,65%. En otras palabras, prácticamente uno de cada dos trabajadores del mercado laboral ecuatoriano lo hace al margen de la regulación laboral, tributaria o de seguridad social. Sin embargo, esta cifra se eleva aún más si se incorpora un segmento poblacional que la metodología oficial invisibiliza: el empleo doméstico sin afiliación a la seguridad social. En efecto, 163.817 trabajadoras domésticas se encuentran en esta situación, lo que elevaría la tasa de informalidad al 52%. La exclusión metodológica de este grupo plantea una interrogante metodológica que merece atención: ¿qué justifica que la definición oficial excluya al servicio doméstico no afiliado del cómputo de la informalidad, si conceptualmente pertenece a la economía transaccional y no al autoconsumo del hogar? Esta decisión de diseño muestral, cuyas razones técnicas no se documentan públicamente con suficiente detalle, tiene una consecuencia práctica considerable: incorporando este segmento, la tasa de informalidad no solo se eleva a 52%, sino que el discurso cambia para empezar a hablar de que en 2025 la mayoría de los trabajadores ecuatorianos labora en la informalidad.

Desde el punto de vista técnico, el empleo doméstico no puede considerarse autoconsumo, sino que forma parte de las transacciones de consumo de los hogares en la economía. Por lo tanto, debería incluirse en la contabilización de la informalidad, ya que constituye un intercambio económico con implicaciones fiscales y laborales claras. La omisión de este segmento no solo distorsiona la medición, sino que también invisibiliza a un grupo históricamente vulnerable, mayoritariamente compuesto por mujeres, que enfrenta una precariedad laboral estructural.

Esta cifra, incluso en su versión conservadora, desmiente cualquier lectura de la informalidad como un fenómeno residual o marginal: se trata de un régimen laboral que compite en tamaño con la formalidad misma y cuya persistencia revela tanto limitaciones técnicas en la medición como posibles motivaciones políticas en la definición de los criterios oficiales. En este sentido, la comprensión cabal de la magnitud del fenómeno resulta indispensable: sin una medición que refleje con precisión todas sus dimensiones, será difícil evaluar de manera rigurosa las posibilidades de expansión del empleo formal en la economía ecuatoriana, y esta seguirá sin ser una prioridad de política pública.

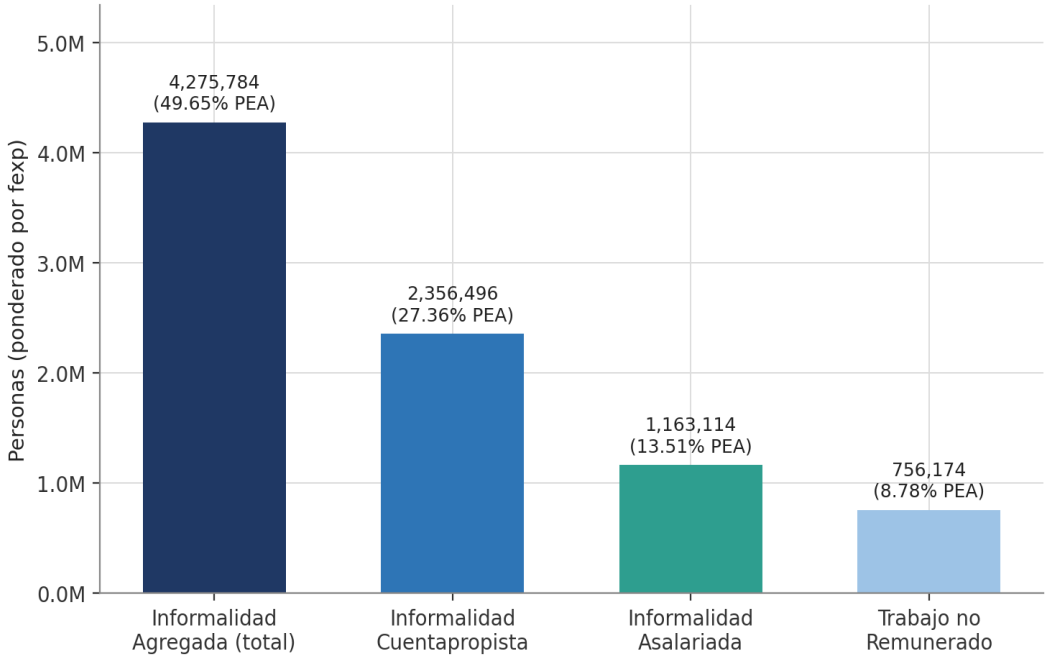


Figura 4. Composición de la informalidad laboral, ponderada por fexp.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto TFM__5_.ipynb, celda 74.

Al desagregar esta magnitud en sus componentes internos, emerge una segunda lectura, igualmente relevante: la informalidad no constituye un bloque homogéneo, sino un conjunto de tipologías con dinámicas diferenciadas. El segmento predominante es la informalidad cuentapropista, conformada por trabajadores en autoempleo y pequeños emprendimientos de subsistencia, así como por grandes empresarios que operan al margen de la regulación laboral y tributaria. Este grupo concentra 2'356.496 personas, equivalentes al 27,37% de la PEA, lo que lo convierte en el núcleo más numeroso de la informalidad. En segundo lugar, se ubica la informalidad asalariada, integrada por trabajadores que dependen de un empleador pero carecen de estabilidad contractual y de acceso a la protección social, debido a que este empleador no cuenta con RUC ni registros contables, por lo que no puede afiliar a su fuerza laboral ni ofrecerles estabilidad contractual. Este segmento alcanza 1'163.114 personas, es decir, el 13,51% de la PEA. Finalmente, el trabajo no remunerado, compuesto por personas que realizan

actividades productivas sin recibir ingresos directos, representa 756.174 personas, equivalentes al 8,78% de la PEA.

Esta desagregación permite observar con claridad que la informalidad no se reduce únicamente al autoempleo, como suele asumirse en lecturas simplistas, sino que también incluye formas asalariadas y no remuneradas en pequeñas, medianas y grandes unidades productivas, lo que amplía su alcance y complejidad. El Cuadro 2 resume esta composición:

Categoría	Personas (ponderado)	% de la PEA
Informalidad Agregada	4.275.784	49,65%
Informalidad Cuentapropista	2.356.496	27,37%
Informalidad Asalariada	1.163.114	13,51%
Trabajo no Remunerado	756.174	8,78%
PEA total (referencia)	8.612.073	100%

Cuadro 2. Composición de la informalidad laboral (valores ponderados por fexp).
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 74.

El hecho de que la informalidad cuentapropista supere en magnitud a la informalidad asalariada ofrece una lectura relevante para el análisis. Esta composición indica que una parte sustancial del fenómeno no puede abordarse únicamente mediante la exigencia de cumplimiento normativo a empleadores identificables, sino que requiere atender las condiciones de productividad, de acceso a mercados y de sostenibilidad de miles de unidades de autoempleo dispersas en todo el territorio nacional. En este segmento, la informalidad se vincula más con la capacidad de generar ingresos en contextos de baja capitalización y limitada inserción en cadenas de valor, que con la evasión deliberada de obligaciones legales. Atender este segmento en función de su tamaño, exige que los hacedores de política laboral, tributaria y fiscal prioricen dentro de sus políticas públicas mecanismos de seguimiento, control e incentivos que promuevan la formalización de estas unidades productivas y de las relaciones laborales que al interior de estas coexisten, alcanzando empresas medianas y grandes.

Finalmente, el trabajo no remunerado, aunque representa el componente de menor magnitud relativa, sigue siendo significativo: casi ocho de cada cien integrantes de la PEA participan en actividades productivas sin percibir remuneración directa. Estas actividades suelen desarrollarse en negocios familiares de comercio, manufactura o alojamiento, o en parcelas agrícolas, y aunque no generan ingresos monetarios, sí contribuyen al funcionamiento económico de los hogares y a la reproducción de la economía popular. Dentro de este grupo también se incluye el voluntariado, que no necesariamente implica explotación laboral, sino que puede responder a una decisión altruista de contribuir al bienestar colectivo. No obstante, es importante reconocer que en ciertos sectores empresariales esta modalidad puede convertirse en un mecanismo para abaratar costos, como ocurre en actividades culturales, deportivas o de servicios comunitarios, donde las organizaciones recurren sistemáticamente a voluntarios para cubrir funciones que en otros contextos serían remuneradas.

Este doble carácter, altruista en algunos casos y funcional a la reducción de costos en otros, plantea la necesidad de una regulación específica que permita distinguir entre el voluntariado genuino y las prácticas empresariales que lo utilizan como sustituto del empleo formal. De este modo, se incentivaría que el trabajo no remunerado fuera reconocido por su aporte social y comunitario, sin convertirse en un mecanismo de precarización laboral encubierta.

4.1.2. Un país de dos velocidades: la brecha urbano-rural

En términos agregados, al igual que en diversas problemáticas laborales, la informalidad oculta una fractura territorial significativa. En el área urbana, se estima 2'939.932 trabajadores formales en 2025 frente a 2'161.384 informales, 178.878 personas en servicio doméstico y 68.365 no clasificadas, lo que sitúa la tasa de informalidad total urbana (informal más servicio doméstico) en 43,75%. En el área rural, el panorama se invierte drásticamente: los 2'114.401 trabajadores informales superan por más de tres veces a los 714.503 formales, y 42.073 personas adicionales están en servicio doméstico, lo que eleva la tasa de informalidad rural a un crítico 72,95%, cifra que casi duplica la observada en las ciudades.

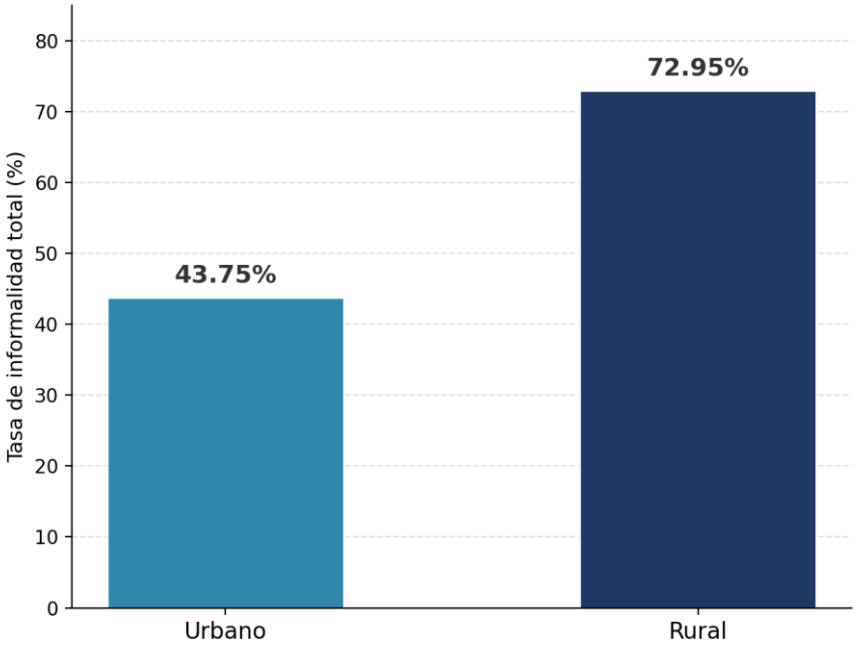


Figura 5. Tasa de informalidad laboral total (informal + servicio doméstico) por área.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celdas 77-78.

Área	Formal	Informal	Serv. doméstico	No clasificados	Informalidad total
Urbano	2.939.932 (54,97%)	2.161.384 (40,41%)	178.878 (3,34%)	68.365 (1,28%)	43,75%
Rural	714.503 (24,17%)	2.114.401 (71,53%)	42.073 (1,42%)	84.986 (2,88%)	72,95%

Cuadro 3. Distribución de la población ocupada por área y categoría laboral (valores ponderados; % por fila).

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celdas 77-78.

Esta brecha no es casualidad, sino el resultado de limitaciones estructurales históricas. Perry et al. (2007) sostienen que en las zonas rurales la informalidad tiende a asociarse con la baja productividad de las actividades agropecuarias de subsistencia y con la débil capacidad institucional para fiscalizar unidades productivas dispersas y de pequeña escala. En el ámbito urbano, en cambio, la persistencia de la informalidad, pese a una mayor densidad de empresas formales, suele explicarse por la urbanización acelerada y la incapacidad del sector moderno para absorber el flujo migratorio proveniente del campo. Bajo esta óptica, la informalidad urbana funciona como un mecanismo de refugio para quienes cuentan con capital humano limitado y, ante la escasez de vacantes formales, optan por el autoempleo o el trabajo en microunidades no reguladas; mientras que en el sector rural la informalidad es prácticamente una condición intrínseca a la estructura productiva dominante.

No obstante, una mirada más reciente revela que la informalidad rural no se explica únicamente por la baja productividad agropecuaria. En 2026, persisten factores adicionales que refuerzan esta condición: la escasez de empleo asalariado estable en territorios rurales, la segmentación estacional de la demanda laboral (por ejemplo, las cosechas y los ciclos agrícolas) y la pluriactividad de los hogares, que combinan la agricultura de subsistencia con actividades informales no agrícolas, como el comercio local, la construcción o los servicios comunitarios. A ello se suma el déficit de infraestructura y de conectividad digital, que limita la integración de productores rurales a las cadenas de valor formales y restringe su acceso a servicios financieros, a la capacitación y a mercados más amplios.

En este sentido, la informalidad rural debe entenderse como el resultado de una estructura productiva fragmentada, pero también de condiciones contemporáneas particularmente adversas que perpetúan la precariedad: baja densidad de empleadores formales, ausencia de políticas de empleo rural sostenido y limitaciones tecnológicas que impiden aprovechar plenamente la digitalización. Así, mientras la informalidad urbana se transforma con fenómenos como la gig economy, la ruralidad sigue marcada por la falta de alternativas asalariadas y por la reproducción de estrategias de subsistencia que mantienen a la mayoría de los trabajadores fuera de la regulación laboral.

4.1.3. Género e informalidad: un patrón más matizado de lo esperado

A nivel nacional, la tasa de informalidad es prácticamente equivalente entre hombres y mujeres, con una leve diferencia a favor de los hombres, es decir, una tasa apenas mayor entre ellos: 50,74% para los hombres frente a 48,17% para las mujeres. Sin embargo, este promedio agregado oculta una composición territorial más matizada, que solo se revela al desagregar simultáneamente por género y área de residencia.

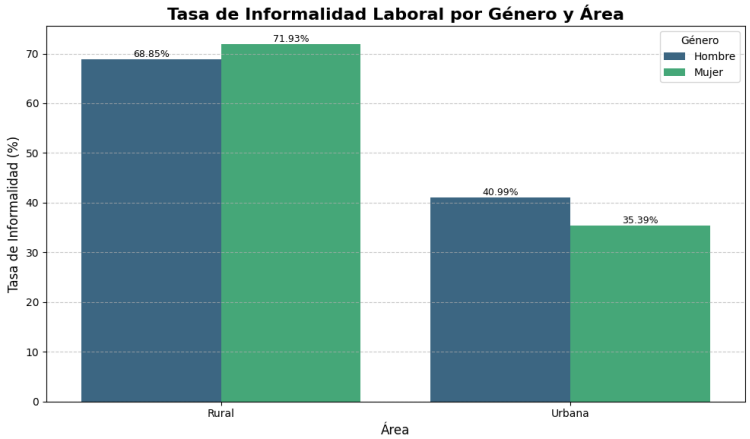


Figura 6. Tasa de informalidad laboral por género y área.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 86.

Grupo	Tasa de informalidad
Hombre (nacional)	50,74%
Mujer (nacional)	48,17%
Hombre — Rural	68,85%
Mujer — Rural	71,93%
Hombre — Urbano	40,99%
Mujer — Urbano	35,39%

Cuadro 4. Tasa de informalidad laboral por género y área.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celdas 85-86.

En el área rural, las mujeres presentan una tasa de informalidad superior a la de los hombres (71,93% frente a 68,85%), lo cual es consistente con la literatura sobre la segregación ocupacional y las responsabilidades de cuidado no remunerado que limitan el acceso de las mujeres a los escasos empleos formales disponibles en el campo. En el área urbana, en cambio, el patrón se invierte: los hombres registran una tasa de informalidad mayor que la de las mujeres (40,99% frente a 35,39%). Una explicación plausible, coherente con la definición de las categorías utilizadas en este estudio, es que una porción importante del empleo informal femenino urbano se concentra en el servicio doméstico, categoría que la ENEMDU contabiliza de forma separada de la "informalidad" propiamente dicha; de incorporarse dicha categoría, la brecha de género en el área urbana probablemente se reduciría o se revertiría. Esta composición diferenciada, y no una menor exposición femenina al trabajo precario, explicaría que el promedio nacional muestre a los hombres con una tasa levemente superior a la de las mujeres.

4.1.4. La heterogeneidad sectorial de la informalidad

La informalidad tampoco se distribuye de manera uniforme entre las ramas de la actividad económica. Al desagregar a los trabajadores formales e informales (asalariados, cuentapropistas y no remunerados) por rama de actividad, se observa que Agricultura, Ganadería y Pesca concentra el mayor volumen de informalidad del país, dominado por el cuentapropismo de subsistencia y el trabajo no remunerado. En contraste, sectores como el comercio presentan una alta densidad de informales asalariados, reflejo de la baja barrera de entrada y de la proliferación de microunidades de venta y servicios que operan sin regulación. Por su parte, Alojamiento y Servicios de Comida se caracteriza por unidades productivas pequeñas, con alta rotación laboral y, más recientemente, por la expansión de emprendimientos vinculados a plataformas digitales de entrega y reservas, que reproducen dinámicas de precariedad laboral bajo esquemas flexibles.

Harris y Todaro (1970) explicaban que estos sectores actúan como receptores del excedente de mano de obra que el sector moderno no logra absorber, mientras que la CEPAL (2020) los identifica como actividades de "baja productividad", en las que la informalidad no responde a una elección individual sino a una estrategia de supervivencia frente al dualismo estructural entre precarización y desempleo, característico de las economías latinoamericanas. Una lectura más apegada a las dinámicas laborales de 2026 permite comprender que estas ramas también funcionan como espacios de innovación informal, donde emergen nuevas dinámicas como el comercio electrónico de baja escala, el e-commerce autónomo, el turismo comunitario y los servicios de entrega, que generan ingresos, pero lo hacen bajo esquemas laborales sin protección social ni estabilidad contractual. El retail, por ejemplo, en su versión informal, se expresa en microunidades de venta sin RUC, sin registros contables y con trabajadores sin afiliación a la seguridad social. Esto abarca desde tiendas de barrio no registradas hasta ventas en redes sociales o en plataformas digitales sin regulación.

En el extremo opuesto, sectores como Administración Pública y Defensa, Enseñanza y Actividades Financieras y de Seguros presentan las tasas de formalización más altas del país. Según Levy (2008), estos sectores operan bajo un marco de alta productividad y estrecha vigilancia regulatoria, donde el costo de operar dentro de la ley se ve compensado por economías de escala y el acceso al crédito, lo que confirma la tesis de Tokman (2001) sobre la segmentación del mercado laboral: existen sectores protegidos por auténticas "barreras de entrada" institucionales que impiden la precarización propia del segmento informal. En la actualidad, además, estos sectores se benefician de la digitalización avanzada, la automatización de procesos y la creciente demanda de servicios especializados, lo que refuerza su capacidad para mantener altos niveles de formalización y absorber trabajadores con mayor capital humano.

En conjunto, esta heterogeneidad sectorial sugiere que la informalidad ecuatoriana tiene un componente estructural-productivo que exige políticas diferenciadas: por un lado, incentivos a la productividad y la digitalización en sectores de baja escala, y por otro, el fortalecimiento de la inspección laboral y la regulación de plataformas digitales en sectores mixtos como el comercio y los servicios. Esta lectura permite entender la informalidad no solo como un vestigio histórico de baja productividad heredado por el mercado laboral, sino como un fenómeno en transformación latente, que incorpora nuevas dinámicas de mercado y tecnologías y que, por ello, requiere respuestas adaptadas a las realidades del siglo XXI.

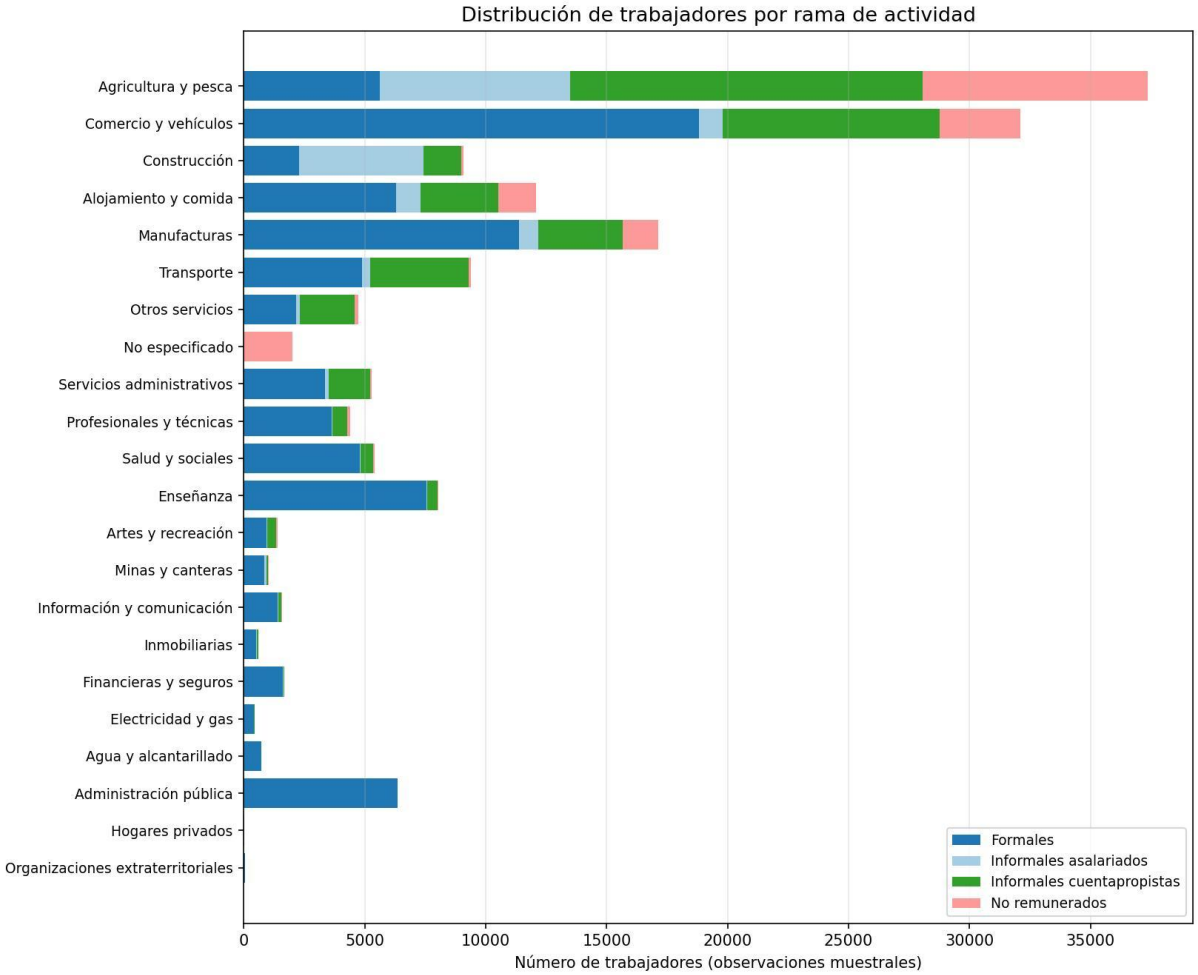


Figura 6.1. Tasa de informalidad laboral por rama de actividad.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 77. Nota: la visualización original se generó con Bokeh como gráfico interactivo; la imagen incluida corresponde a una captura de pantalla de dicha visualización.

4.1.5. El ciclo vital de la informalidad: una trayectoria ascendente tras la juventud

La relación entre edad e informalidad es un elemento de análisis relevante relevante para comprender las dinámicas del mercado laboral en 2026: Si la informalidad fuera principalmente un reflejo de productividad, se esperaría que se concentrara en los años de mayor rendimiento económico de las personas; en cambio, si se tratara en realidad del reflejo de la vulnerabilidad y ausencia de medios de vida de la población, debería ser más alta en los extremos etarios: jóvenes en proceso de inserción y adultos mayores en riesgo de exclusión. Explorar esta relación permite, por tanto, contrastar dos hipótesis sobre el significado de la informalidad en la trayectoria laboral y aportar evidencia sobre si se comporta como un fenómeno de eficiencia o de fragilidad social.

En este sentido, se observa que la relación entre la informalidad y la edad adopta una curva en forma de U, lo que aporta evidencia de la naturaleza dual de la informalidad como fenómeno de vulnerabilidad en los extremos del ciclo laboral. Los jóvenes de 15 a 19 años registran una tasa de informalidad elevada, de 66,57%, reflejo de una inserción laboral temprana y precaria, marcada por empleos de baja calidad y ausencia de protección social. Conforme se acumula experiencia y formación, la tasa desciende de

manera sostenida hasta alcanzar su punto más bajo entre los 25 y 29 años (36,77%), edad en la que se concentra la mayor formalización del empleo y se materializa el retorno de la inversión en capital humano.

A partir de los 30-34 años (38,84%), la tasa vuelve a incrementarse de manera sostenida y prácticamente monotónica en cada tramo etario adicional, hasta alcanzar el 64,80% en el grupo de 60 a 64 años y un nivel crítico del 89,62% en el grupo de 80 a 84 años, el último tramo con observaciones suficientes en la muestra. Esta tendencia ascendente en edades avanzadas refleja la pérdida de empleabilidad en el sector formal, la transición hacia ocupaciones de subsistencia y, en muchos casos, la reinserción en actividades informales como estrategia de supervivencia económica.

La forma en U de la curva evidencia que la informalidad no es un reflejo de un correlato de productividad; lo sería si se concentrara en edades intermedias. Sino más bien constituye un reflejo de vulnerabilidad en los extremos del ciclo vital: jóvenes que ingresan al mercado laboral sin credenciales suficientes y adultos mayores que enfrentan la exclusión del empleo formal.

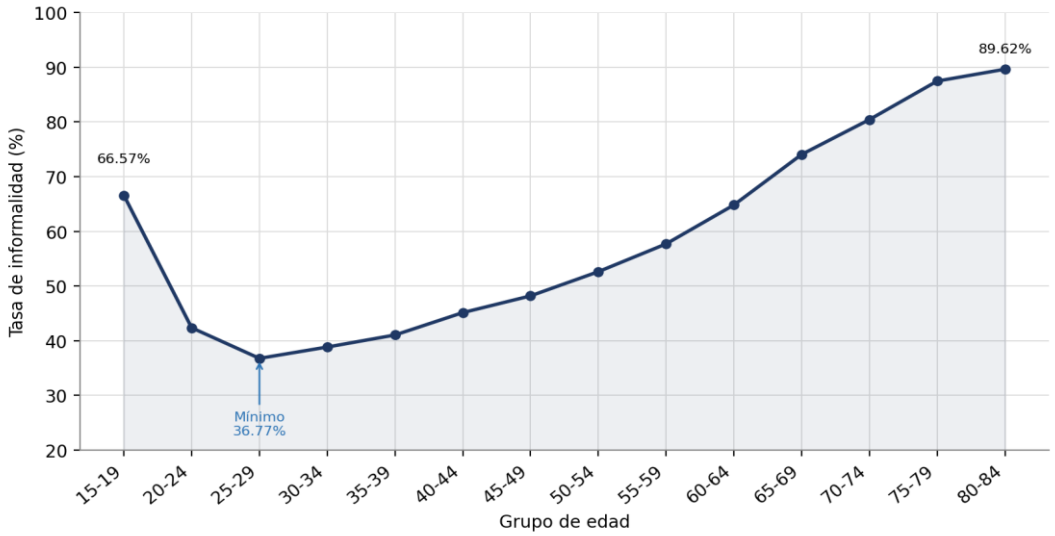


Figura 7. Tasa de informalidad laboral por grupo de edad.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 89.

Grupo de edad	Tasa	Grupo de edad	Tasa
15-19	66,57%	50-54	52,61%
20-24	42,36%	55-59	57,69%
25-29	36,77%	60-64	64,80%
30-34	38,84%	65-69	74,04%
35-39	41,04%	70-74	80,41%

Grupo de edad	Tasa	Grupo de edad	Tasa
40-44	45,13%	75-79	87,45%
45-49	48,19%	80-84	89,62%

Cuadro 5. Tasa de informalidad laboral por grupo de edad (quinquenal).
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 89.

En este contexto, dicha trayectoria tiene una lectura sustantiva: para los más jóvenes, la informalidad opera como la principal, y muchas veces única, puerta de entrada al mercado laboral, aunque a costa de una inserción precaria que dificulta la acumulación temprana de capital humano formal. A partir de la mediana edad, en cambio, la informalidad se convierte en la alternativa que absorbe a los trabajadores que la discriminación por edad, la obsolescencia de competencias y el acceso insuficiente a la afiliación social, van expulsando del sector formal. En consecuencia, la distribución en forma de U de la informalidad según rango etario, lejos de ser una puerta de entrada a medios de subsistencia deseables y adecuados, se perpetúa en ambos extremos del ciclo vital como una relación laboral de baja calidad, lo que exige políticas específicas tanto para facilitar la primera inserción formal de los jóvenes como para sostener la permanencia de los trabajadores de mayor edad en el mercado regulado.

4.1.6. La educación como escudo: el gradiente educativo inverso

El nivel de instrucción es, de los determinantes analizados hasta aquí, el que presenta la relación más nítida y monotónica con la informalidad. Entre quienes no cuentan con ningún nivel de instrucción, la tasa de informalidad alcanza el 87,45% y llega al 93,85% entre quienes han asistido únicamente a un centro de alfabetización. Quienes completaron primaria o educación básica mantienen tasas igualmente elevadas: 71,79% y 71,85%, respectivamente. La tasa desciende de forma notable entre quienes alcanzaron secundaria (45,96%) o bachillerato (44,34%), y cae de manera pronunciada entre quienes cursaron educación superior no universitaria (20,19%) o universitaria (20,05%). En el extremo superior, las personas con posgrado registran una tasa de informalidad de apenas 3,71%.

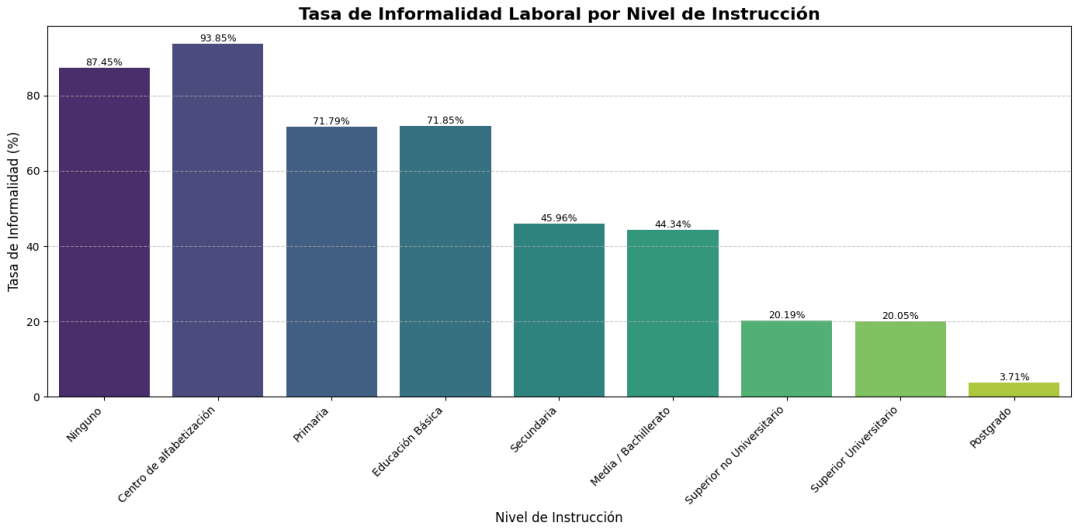


Figura 8. Tasa de informalidad laboral por nivel de instrucción.

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 92.

Nivel de instrucción	Tasa de informalidad
Centro de alfabetización	93,85%
Ninguno	87,45%
Educación Básica	71,85%
Primaria	71,79%
Secundaria	45,96%
Media / Bachillerato	44,34%
Superior no Universitario	20,19%
Superior Universitario	20,05%
Postgrado	3,71%

Cuadro 6. Tasa de informalidad laboral por nivel de instrucción.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 92.

Este gradiente inverso confirma a la educación como uno de los mecanismos de protección más efectivos frente a la informalidad. Levy y Székely (2016) documentan un patrón similar para la región: las generaciones con más años de escolaridad exhiben sistemáticamente tasas de informalidad más bajas, mientras que organismos como la CEPAL y la OCDE coinciden en que la persistencia de la informalidad en América Latina refleja, en el fondo, limitaciones estructurales del desarrollo económico y social que la sola expansión de la cobertura educativa no basta por sí sola para revertir, pero que sin duda alguna la mitigan de manera sustancial.

4.1.7. La brecha de ingresos: una primera aproximación monetaria

Hasta este punto, la caracterización se ha centrado en la incidencia de la informalidad. Cabe entonces preguntarse cuál es su correlato monetario. La distribución del logaritmo del ingreso laboral evidencia que los trabajadores formales se concentran en niveles de ingreso más altos y con menor dispersión, mientras que los informales exhiben una distribución más dispersa y desplazada hacia ingresos menores, lo que refleja una mayor heterogeneidad y precariedad dentro del grupo.

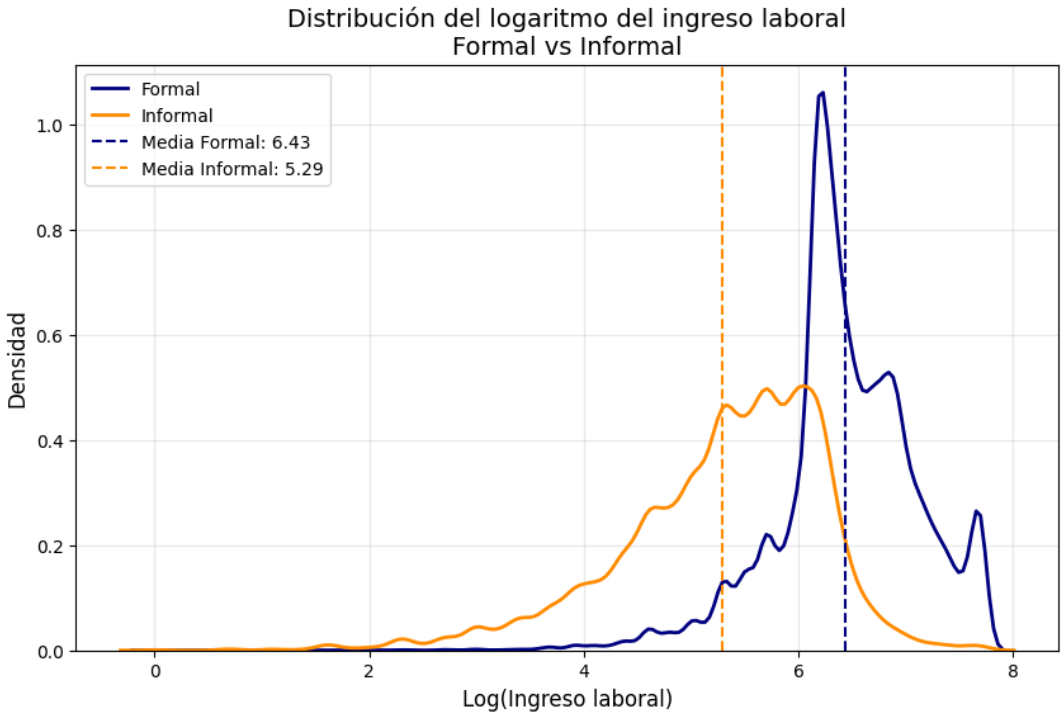


Figura 9. Distribución del logaritmo del ingreso laboral, formal vs. informal.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celdas 95-96.

Condición laboral	Media log(ingreso)	Ingreso promedio ($)
Formal	6,43	621,85
Informal	5,29	198,44

Cuadro 7. Ingreso laboral promedio, formal vs. informal.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celdas 95-96.

Hasta este punto, el análisis se ha centrado en la incidencia de la informalidad desde una perspectiva estructural y sectorial. Un paso siguiente consiste en observar cómo esta condición se refleja en los ingresos laborales, pues la dimensión monetaria ofrece una primera aproximación al impacto económico de la informalidad en la vida de los trabajadores. Examinar esta relación permite identificar si la diferencia entre empleo formal e informal se traduce en una brecha sistemática de ingresos y, en consecuencia, si la informalidad opera como un factor que amplifica las desigualdades en el mercado laboral.

La distribución del logaritmo del ingreso laboral evidencia que los trabajadores formales se concentran en niveles de ingreso más altos y con menor dispersión, lo que refleja una mayor estabilidad y homogeneidad en las remuneraciones. En contraste, los trabajadores informales exhiben una distribución más dispersa y desplazada hacia ingresos menores, signo de una heterogeneidad interna marcada por la coexistencia de ocupaciones de subsistencia, microunidades productivas y empleos precarios sin regulación, o de unidades productivas que, aunque medianas o grandes, deciden operar al

margen de los derechos laborales y de sus obligaciones tributarias y fiscales. Esta diferencia en la forma de la distribución aporta evidencia de que la informalidad no solo implica la ausencia de protección social, sino también una penalización económica directa**.**

En términos concretos, se halla que un trabajador formal percibe en 2025, en promedio, $621,85 mensuales, frente a $198,44 de un trabajador informal. Esto implica que, en promedio, un trabajador formal gana cerca de 3,13 veces el ingreso de un trabajador informal. Esta primera aproximación, aunque agregada, ya deja entrever la magnitud de la penalización monetaria asociada a la informalidad; sin embargo, como se muestra a continuación, dicha brecha no es uniforme, sino que varía sustancialmente según la posición que ocupa cada trabajador dentro de la estructura productiva.

4.1.8. La brecha no es uniforme: penalización salarial por categoría ocupacional

Al desagregar el ingreso promedio ponderado por categoría de ocupación y por condición de formalidad, la brecha salarial se revela heterogénea. Las categorías de Patrono y de Ayudante no remunerado presentan las diferencias más extremas, aunque de naturaleza distinta entre sí. Entre los patronos, la condición informal es prácticamente inexistente en la muestra; el ingreso promedio ponderado formal asciende a $1.122,96, sin un contrapunto informal comparable, de modo que la brecha no es tanto salarial como de acceso mismo a la condición de empleado (un trabajador informal difícilmente llega a ser patrono), reflejo del capital y las economías de escala que únicamente la estructura formal permite sostener. Entre los ayudantes no remunerados, en cambio, sí existe comparación directa: el ingreso formal ($1.051,45) supera al informal ($536,17) en 96,1%, una brecha que prácticamente duplica el ingreso de quien trabaja en condición formal; sin embargo, este salario en realidad estaría reflejando el número de horas disponibles para una ocupación o trabajo secundario del que los ayudantes no remunerados sí pueden obtener un ingreso; pues por definición su ocupación primaria no les permite obtener una remuneración. Entre los trabajadores por cuenta propia, la brecha también es sustancial: - un cuentapropista formal percibe $601,44 frente a $313,39 de su par informal, un 91,9% más.

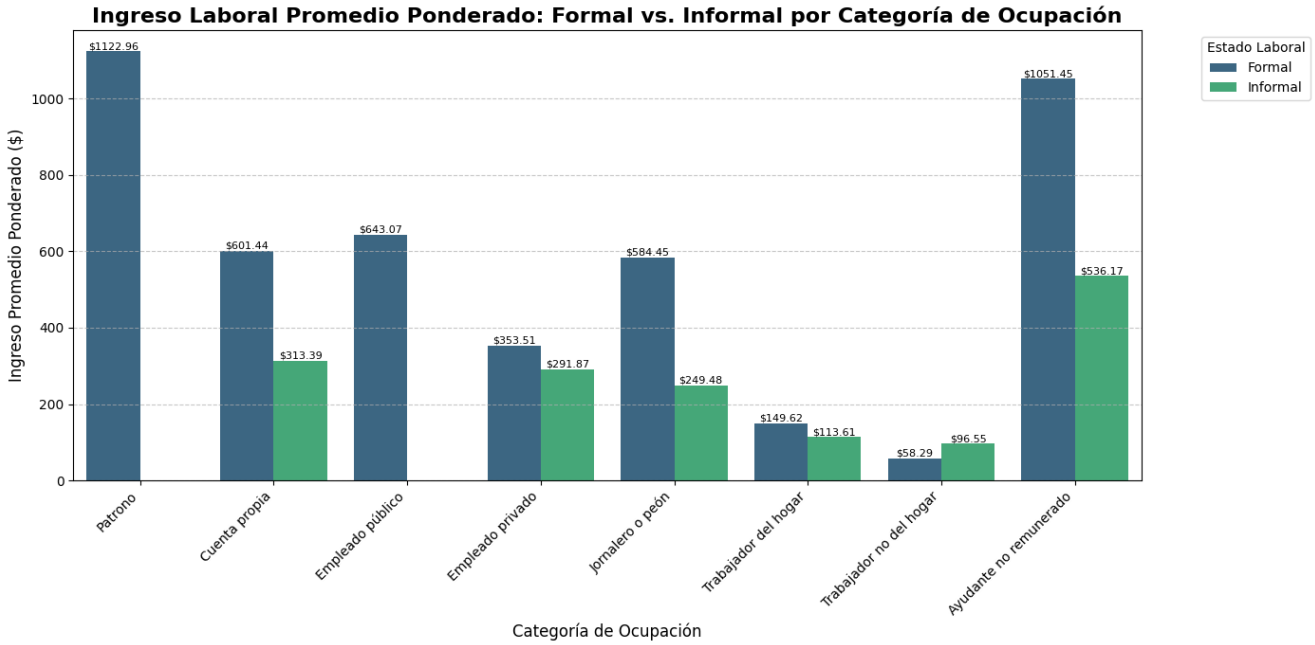


Figura 10. Ingreso laboral promedio ponderado, formal vs. informal, por categoría de ocupación.
Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 98. Los valores exactos por categoría se muestran como etiquetas sobre cada barra.

En el segmento asalariado (empleados privados y jornaleros o peones), la brecha interna confirma la existencia de lo que puede denominarse una **"**penalización salarial pura**"**: a pesar de desempeñar tareas similares, la ausencia de un marco contractual y de beneficios sociales en la informalidad se traduce directamente en menores ingresos monetarios. Entre empleados privados, el ingreso formal ($353,51) supera al informal ($291,87) en 21,1%; entre jornaleros o peones, la brecha es considerablemente más amplia, con el ingreso formal ($584,45) superando al informal ($249,48) en 134,3%, la mayor penalización porcentual de todo el segmento asalariado.

Si bien Mincer (1974) atribuía estas diferencias principalmente a divergencias en educación y experiencia, estudios más recientes, como los de la CEPAL (2020), muestran que incluso al controlar por el capital humano, la condición de informalidad opera como una barrera estructural que impide la igualación de los retornos salariales. En el contexto actual, esta penalización se explica también por la creciente dualización del mercado laboral: mientras las empresas formales internalizan costos de seguridad social, cumplimiento normativo y capacitación, las unidades informales compiten con márgenes de ganancia reducidos y trasladan la precariedad directamente a los ingresos de sus trabajadores.

La lectura contemporánea del dualismo laboral sugiere que la brecha interna no solo refleja la segmentación productiva, sino también la manera en que la informalidad transiciona a nuevas modalidades. En sectores como la construcción, el comercio y los servicios personales, la informalidad asalariada se ha visto reforzada por la expansión de los contratos temporales, la tercerización y el trabajo mediado por plataformas, que reproducen esquemas de baja protección y salarios inferiores. Así, la penalización salarial pura se convierte en un síntoma de un mercado laboral que combina viejas formas de exclusión con nuevas dinámicas de flexibilización digital.

Los resultados en categorías como la de jornaleros y la de trabajadores del hogar corroboran la tesis de la "informalidad de exclusión": el trabajador no elige la informalidad, sino que es empujado hacia ella ante la falta de vacantes en el sector moderno, y acepta salarios sensiblemente menores. Entre los trabajadores del hogar, por ejemplo, el ingreso formal ($149,62) supera al informal ($113,61) en 31,7%, lo que refleja un segmento históricamente vulnerable y con escasa capacidad de negociación. La categoría de trabajador no del hogar constituye la única excepción a este patrón (el ingreso informal ($96,55) supera al ingreso formal ($58,29) en 65,6%), un resultado que, dado el reducido tamaño de la muestra, debe leerse con cautela antes de generalizarlo.

4.1.9. Trayectorias de ingreso: dualismo estructural y brechas de género

Un aspecto clave para comprender la informalidad es observar cómo evolucionan los ingresos a lo largo del ciclo de vida, diferenciando por género y por condición de formalidad. Este diagnóstico tiene como objetivo identificar patrones que no se captan al analizar cada dimensión por separado y ofrece una visión más completa de las desigualdades presentes en el mercado laboral. Al combinar edad, género y formalidad, se revelan trayectorias que muestran no solo diferencias salariales, sino también las formas en que se segmenta y estratifica el mercado de trabajo, lo que genera brechas persistentes que condicionan las oportunidades de los distintos grupos de trabajadores.

El resultado es el siguiente: Los hombres formales presentan la trayectoria de ingreso más elevada y cóncava a lo largo de todo su ciclo de vida económicamente productivo, alcanzando su punto máximo entre los 40 y 50 años de edad; lo cual es un patrón consistente con la acumulación de experiencia predicha por la teoría del capital humano. En el extremo opuesto, las mujeres informales se ubican de manera sostenida en el estrato más bajo del gráfico, con una curva casi plana que evidencia el

estancamiento de sus ingresos independientemente de la edad o la experiencia acumulada; así como una brecha salarial sostenida respecto a todos los grupos con los que se comparan.

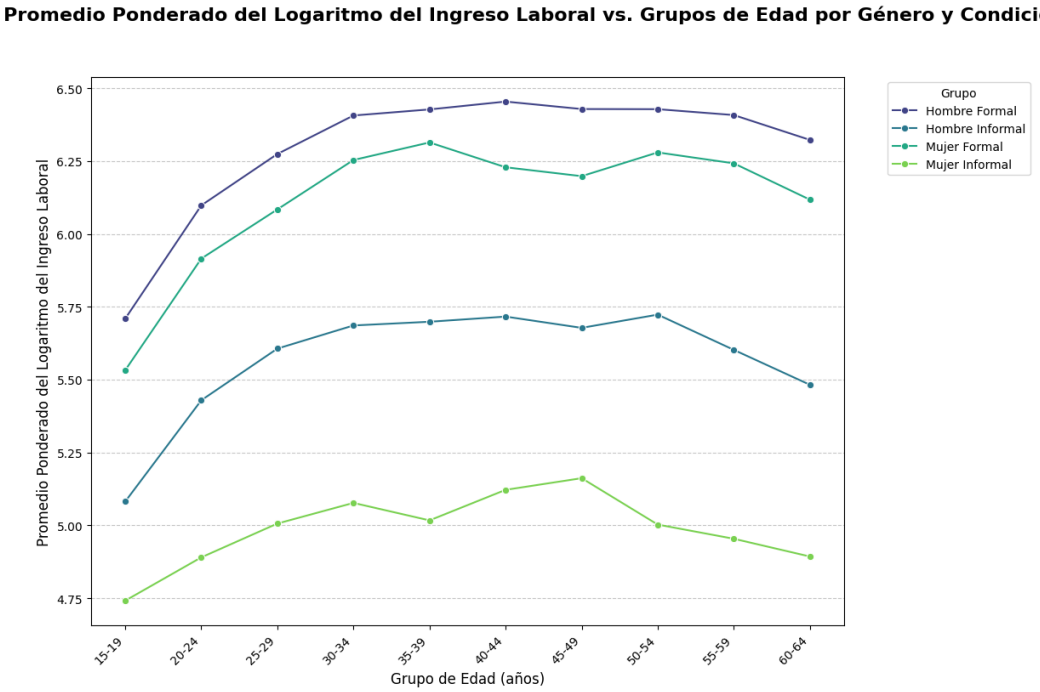


Figura 11. Promedio ponderado del logaritmo del ingreso laboral por grupo de edad, género y condición de formalidad.

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, celda 101.

La brecha más severa se observa precisamente al comparar a los hombres formales con las mujeres informales de un mismo rango etario: una disparidad que no es solo monetaria, sino también estructural y refleja una asimetría profunda en el acceso a la protección social y a la estabilidad contractual. Desde la perspectiva minceriana, la horizontalidad de la curva femenina informal sugiere que los retornos a la experiencia son prácticamente nulos en este segmento, lo que invalida la premisa de optimización individual del capital humano y, en cambio, apunta a una falla de mercado. El Banco Interamericano de Desarrollo BID (2015) lo caracteriza mediante la metáfora de "paredes de cristal": barreras invisibles pero rígidas que impiden a la mujer informal transitar hacia sectores de mayor productividad, perpetuando así una brecha de género que la condición de informalidad no hace sino profundizar.

4.1.10. Síntesis: de la caracterización a la estimación

La caracterización desarrollada en las secciones previas muestra que la informalidad ecuatoriana en 2025 es un fenómeno heterogéneo y estructural. Predomina la informalidad cuentapropista por sobre la asalariada, la ruralidad exhibe tasas que duplican o triplican las urbanas, y la relación con la edad adopta una forma en U que convierte a la informalidad en puerta de entrada para los jóvenes y en red de supervivencia para los adultos mayores. A esto se suma la evidencia del gradiente educativo: a mayor nivel de instrucción, menor probabilidad de informalidad. Juntas, estas regularidades no solo describen quiénes son los informales, sino que también señalan por qué las respuestas de política deben ser diferenciadas y territoriales, y no meramente punitivas o uniformes.

Más allá de la descripción, el capítulo aporta una lectura actualizada de las transformaciones que reconfiguran la informalidad: la expansión de las plataformas digitales y del comercio electrónico de

baja escala, la persistente brecha de conectividad en áreas rurales y la coexistencia de microunidades productivas con empresas que operan al margen de la regulación. Estas dinámicas generan formas híbridas de trabajo —innovadoras en su capacidad de generar ingresos, pero precarias en términos de protección social— que requieren instrumentos nuevos: acceso a mercados digitales, simplificación tributaria escalonada y regulación adaptada a plataformas que externalizan riesgos laborales.

La diferencia promedio de ingresos entre formales e informales no es un dato accesorio: es la manifestación monetaria de una desigualdad estructural que erosiona la capacidad de acumulación de activos y la protección social de amplios segmentos de la PEA. Por ello, la priorización de políticas debe orientarse no solo a reducir la incidencia de la informalidad, sino también a mitigar sus efectos distributivos mediante intervenciones que aumenten la productividad, faciliten la transición a la formalidad y protejan a quienes, por edad o por contexto territorial, enfrentan barreras persistentes para acceder al empleo regulado.

Desde el punto de vista metodológico, la sección concluye planteando la necesidad de avanzar desde la asociación descriptiva hacia la estimación causal y la predicción focalizada. Estimar la penalización salarial atribuible estrictamente a la condición de informalidad —controlando por educación, sector, edad, territorio y otras covariables— permitirá cuantificar el impacto real de la informalidad sobre los ingresos. Complementariamente, la construcción de perfiles de riesgo mediante técnicas de machine learning ofrecerá criterios operativos para focalizar programas de formalización y capacitación con mayor eficiencia fiscal y mayor probabilidad de éxito.

En suma, la caracterización aquí presentada funciona como un punto de referencia contextual y como un mapa operativo: identifica los ejes sobre los que deben diseñarse intervenciones diferenciadas (tipología de informalidad, territorio, ciclo vital y nivel educativo) y prepara el terreno para las secciones siguientes, que buscan transformar ese mapa en evidencia causal y en herramientas predictivas. En este sentido, la contribución de esta sección no es solo diagnosticar un problema conocido, sino ofrecer una base empírica y analítica que permita priorizar políticas que reduzcan la informalidad y atenúen sus efectos distributivos en el Ecuador contemporáneo.

4.2. Estimación de la Penalización Salarial Asociada a la Informalidad Laboral

La sección anterior estableció que la informalidad laboral ecuatoriana no se distribuye al azar, sino que se concentra en patrones reconocibles según ocupación, territorio, ciclo de vida y género, y que esa concentración tiene una correlación monetaria directa: un trabajador formal percibe, en promedio, 3,13 veces el ingreso de uno informal. Esa cifra, sin embargo, sigue siendo una fotografía cruda. Compara promedios de dos grupos que difieren entre sí en mucho más que su condición de formalidad -en escolaridad, en experiencia, en el sector donde trabajan, en el tamaño del establecimiento que los emplea-, de modo que no permite afirmar cuánto de esa brecha responde estrictamente a la informalidad y cuánto responde, en realidad, al resto de las características socioeconómicas de los trabajadores informales. Aislar esa penalización neta, asociada exclusivamente a la condición de informalidad y que persiste una vez controladas las demás diferencias, es el objetivo de esta sección.

El ejercicio se desarrolla en tres momentos que se superponen, cada uno de los cuales pone a prueba al anterior. Se parte de una ecuación de Mincer clásica, la especificación mínima que la teoría del capital

humano ofrece para explicar el ingreso laboral a partir de la educación y la experiencia, y que se ha aplicado en varias ocasiones en América Latina. A este planteamiento clásico se añade la condición de informalidad como variable explicativa adicional. Se mostrará que esa primera aproximación, sobreestima la penalización al no distinguir entre lo que es efecto puro de la informalidad y lo que responde a la estructura productiva en la que esta se inserta cada trabajador (el tamaño de la firma, la rama de actividad, la posición socioeconómica del hogar), lo que motiva una segunda especificación de la ecuación de Mincer, denominada Mincer extendida, que incorpora esos controles estructurales, para ofrecer una adaptación de las aplicaciones clásicas y latinoamericanas al sistema laboral ecuatoriano. Si bien es cierto que los hallazgos determinarán cuál es la penalización de la informalidad en la vida de las personas, se revisará, complementariamente, cómo las características socioeconómicas y las decisiones de los individuos suponen costos laborales no observados, y de ello se derivarán también importantes elementos de reflexión. El tercer momento recurre a la aplicación y a la descomposición del Modelo Oaxaca-Blinder como contraste metodológico. En lugar de imponer una única función de remuneración para todo el conjunto, el procedimiento estima por separado una ecuación de salarios para los trabajadores formales y otra para los informales; a partir de esas estimaciones se descompone la brecha salarial observada en dos componentes principales: uno "explicado" por diferencias en dotaciones observables (educación, experiencia, sector, tamaño de la empresa, etc.), y otro "no explicado" que recoge la distinta remuneración que el mercado asigna a esas mismas dotaciones según la condición de formalidad. Técnicamente, la parte explicada refleja qué tanto de la brecha se debe a que los grupos tienen características socioeconómicas distintas, mientras que la parte no explicada captura el efecto de remuneración o de discriminación salarial asociado exclusivamente a la condición laboral.

En consecuencia, esta sección ofrece, además, una oportunidad metodológica que trasciende el hallazgo sustantivo. La descomposición de Oaxaca-Blinder se implementa aquí no solo con las herramientas propias de la econometría clásica, sino también incorporando una lógica de validación cruzada —la separación entre conjuntos de entrenamiento y de prueba— más habitual en la ciencia de datos que en la práctica econométrica convencional. Ese cruce no es incidental: como argumentan Varian (2014) y Mullainathan y Spiess (2017), buena parte del valor que la ciencia de datos puede aportar a las ciencias sociales no reside en sustituir sus técnicas tradicionales, sino en dotarlas de protocolos de validación más exigentes que evalúen su capacidad de generalizar más allá de la muestra con la que fueron estimadas. En ese sentido, esta sección no solo busca cuantificar la penalización salarial de la informalidad en Ecuador, sino también ilustrar, en un caso concreto, cómo la econometría y el aprendizaje automático pueden complementarse en lugar de competir para generar insights más robustos.

4.2.1. El punto de partida: la insuficiencia de la ecuación de Mincer clásica

Toda estimación de la penalización salarial de la informalidad necesita, antes que nada, un modelo capaz de explicar por qué dos personas con características observables distintas terminan percibiendo ingresos distintos. Ese es, precisamente, el problema que la ecuación de Mincer busca resolver, y por esa razón constituye el punto de partida natural de esta sección: no necesariamente porque sea la técnica más sofisticada disponible, sino porque es la más transparente y sobre la que académicamente existe una convención para aislar, paso a paso, cuánto de la brecha salarial responde al capital humano de cada trabajador y cuánto corresponde, una vez controlado ese capital humano, a su condición de informalidad. Esta primera subsección presenta la especificación más simple de ese ejercicio, explica su lógica y sus componentes, estima sus coeficientes sobre la población ecuatoriana ocupada y, sobre esa base,

identifica qué queda todavía sin explicar y por qué, lo que motiva la especificación extendida que se desarrolla en la subsección siguiente.

La ecuación de Mincer, formulada por Jacob Mincer (1974) a partir de la teoría del capital humano, propone que el logaritmo del ingreso laboral de un trabajador puede explicarse, en buena parte, por dos inversiones que esa persona ha hecho a lo largo de su vida: los años que dedicó a estudiar y los que dedicó a trabajar. La intuición detrás de la especificación logarítmica, y no del ingreso expresado en niveles, es igualmente sencilla: la teoría del capital humano predice que cada año adicional de estudio o de experiencia incrementa el ingreso en un porcentaje aproximadamente constante, no en un monto fijo de dólares, de modo que es el logaritmo del ingreso, y no el ingreso mismo, el que guarda una relación lineal con esas dos inversiones. Pese a haberse formulado hace más de cinco décadas, la especificación conserva una vigencia notable en la investigación empírica contemporánea: Lemieux (2006), en su revisión de treinta años de aplicaciones de la ecuación, observa que la inmensa mayoría de los estudios sobre determinación salarial que se publican en la actualidad siguen ajustándose, en sus elementos centrales, a la misma forma funcional que Mincer propuso originalmente, lo que la ha convertido en un lenguaje común que permite comparar resultados entre países, períodos y poblaciones distintas, incluida, para los fines de este estudio, la comparación entre trabajadores formales e informales en Ecuador.

Antes de presentar su forma funcional, conviene precisar qué mide cada uno de sus componentes, porque es sobre esas definiciones en las que se apoya toda la interpretación posterior. El logaritmo del ingreso laboral, abreviado como log_ingrl, es la variable que el modelo busca explicar. La escolaridad, abreviada como year_sh, mide los años de educación formal completados por el trabajador. La experiencia potencial, abreviada exp_sh, no es un dato que la encuesta releve directamente, sino una construcción estándar de la literatura: se calcula como la edad del trabajador menos sus años de escolaridad menos seis años adicionales, estos últimos en representación de la edad a la que, en el sistema educativo ecuatoriano, un niño ingresa típicamente a la educación formal. Restar la escolaridad y esos seis años a la edad es, en otras palabras, una manera de aproximar cuántos años ha tenido esa persona disponibles para trabajar desde que, en teoría, podría haber salido del sistema educativo. Junto a la experiencia potencial se incluye su término elevado al cuadrado, exp_sh^2, no como un simple recurso estadístico sino porque la teoría del capital humano predice que el ingreso no crece de forma indefinida con la experiencia: los primeros años de vida laboral rinden más, en términos de ingreso adicional, que los años posteriores, de modo que el perfil de ingresos a lo largo de la vida de un trabajador describe una curva cóncava, que asciende con fuerza al inicio y se aplana -e incluso declina- hacia el final de la vida laboral. El término cuadrático es, precisamente, el que permite que la ecuación capture esa curvatura: sin él, el modelo forzaría una relación estrictamente creciente entre experiencia e ingreso, algo que la evidencia empírica contradice. A estas dos variables centrales de la teoría del capital humano se añaden, para los fines específicos de esta investigación, la condición de informalidad del trabajador (variable dicotómica que se abrevia es_informal, con valor de uno si el trabajador es informal y de cero si es formal, y cuyo coeficiente constituye el objeto central de esta sección) y su área de residencia, otra variable dicotómica, rural, que distingue entre trabajadores urbanos y rurales. Con estos elementos, la ecuación de Mincer clásica que se estima en esta sección adopta la siguiente forma:

$$log_ingrl = \beta_0 + \beta_1 \cdot es_informal + \beta_2 \cdot year_sh + \beta_3 \cdot exp_sh + \beta_4 \cdot exp_sh^2 + \beta_5 \cdot rural + \varepsilon$$

donde cada coeficiente β mide el efecto marginal de su variable correspondiente sobre el logaritmo del ingreso laboral, manteniendo constante el resto de variables del modelo, y ε recoge la porción del ingreso que la ecuación no logra explicar.

Con este planteamiento general, antes de reportar los resultados, conviene explicitar con precisión qué se hizo. La ecuación se ajustó por mínimos cuadrados ordinarios sobre 96.311 trabajadores ocupados, seleccionados aleatoriamente para conformar el 70% de la muestra objetivo destinada a la estimación de coeficientes; el 30% restante, reservado como conjunto de prueba, permite verificar más adelante que el modelo generaliza fuera de la muestra con la que fue entrenado. El Cuadro 8 reporta, para cada variable, su coeficiente estimado, el error estándar asociado, el estadístico t y su significancia, y el intervalo de confianza al 95%, en el mismo formato en que se presentan los resultados de todos los modelos a lo largo de este capítulo.

Cuadro 8. Ecuación de Mincer clásica (conjunto de entrenamiento).

R^2 = 0,377 | R^2 ajustado = 0,377 | Estadístico F = 1,164e+04 | N (entrenamiento) = 96,311

Variable	**Coef.**	**Error est.**	**t**	**P>|t|**	**IC 95%**
Constante	5,5281	0,012	444,417	<0,001	[5,504, 5,552]
Informalidad	-0,7983	0,005	-149,895	<0,001	[-0,809, -0,788]
Escolaridad (year_sh)	0,0388	0,001	57,724	<0,001	[0,038, 0,040]
Experiencia potencial (exp_sh)	0,0290	0,000	67,210	<0,001	[0,028, 0,030]
Experiencia potencial² (exp_sh²)	-0,0005	6,84e-06	-71,336	<0,001	[-0,001, -0,000]
Área rural	-0,0939	0,006	-16,606	<0,001	[-0,105, -0,083]

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.1.

La lectura de estos coeficientes sigue un orden deliberado: primero, el que constituye el objetivo central de esta sección, la informalidad; luego, los dos que conforman el núcleo de la teoría del capital humano, escolaridad y experiencia; y, finalmente, el que se añadió como control adicional mediante el área de residencia.

En este sentido, se observa que el coeficiente de informalidad es negativo (β = -0,7983) y estadísticamente significativo al 1%. Su magnitud, expresada en unidades logarítmicas, no es directamente intuitiva, por lo que su interpretabilidad requiere una traducción a una magnitud relativa (brecha) o a una monetaria (USD). Transformando matemáticamente este coeficiente logarítmico, se obtiene que, en promedio, un trabajador informal percibe apenas el 45,0% del ingreso de un trabajador formal con la misma escolaridad, la misma experiencia y la misma área de residencia, o lo que es lo mismo, un trabajador informal percibe un 55,0% menos que el salario de un trabajador formal. Esta cifra es, en una primera aproximación válida, sobre cuánto le cuesta a un trabajador ecuatoriano encontrarse en situación de informalidad. Para ponerlo en perspectiva, si un trabajador formal en el país con un nivel de instrucción básica como máximo nivel educativo alcanzado gana USD 482 al mes, es decir, logra cubrir el salario básico unificado, un trabajador informal equivalente con el mismo nivel de educación formal ganará 217 USD al mes, con el consiguiente déficit que ello implica, tanto en consumo vital

como en consumo suntuario. Por ello, para llegar al menos al salario básico, deberá ofertar más horas de trabajo (alternativa poco viable, pues el desempleo es una de las causas de la informalidad, como se vio en los fundamentos teóricos de este estudio), o recurrir a financiamiento que le permita cubrir dicho consumo vital; e incluso deberá considerar incorporar a nuevos miembros de su hogar a la fuerza laboral. De hecho en el país, el tamaño de hogar de jefes de hogares con trabajo informal es ligeramente superior al tamaño de jefes de hogar con trabajo formal (3,36 miembros, en promedio vs. 3,32 miembros en promedio, respectivamente); siendo esta un diferencia estadísticamente significativa. Sin embargo, como se explica al final de esta sección, la brecha estimada bajo el enfoque clásico de Mincer todavía mezcla el efecto propio de la informalidad con el de otras características en las que informales y formales tienden a diferir de manera sistemática.

Continuando con la interpretación de los coeficientes, ambos coeficientes de capital humano se comportan exactamente como la teoría los anticipa. La escolaridad ($\beta = 0{,}0388$) es positiva y significativa: cada año adicional de educación formal se asocia con un incremento de 4,0% en el ingreso laboral esperado, manteniendo constantes la experiencia, la informalidad y el área de residencia. Por ejemplo:

Si se analiza al mismo trabajador (formal, urbano, 10 años de experiencia) con 9 vs. 10 años de escolaridad, el primero ganaría $453,40 y el segundo $471,39. Un 3,97% de penalización salarial (prácticamente el 4,0% del coeficiente). De la misma manera, un trabajador con primaria completa (6 años) vs. Un trabajador con bachillerato completo (12 años) y el mismo tiempo de experiencia, recibe una remuneración mensual $403,42 frente a $509,55 que podría recibir si habría completado su bachillerato. Se puede, incluso con fines ilustrativos, añadir un tercer perfil: con diecisiete años de educación formal (estudios universitarios completos) y treinta y tres años de edad. Su ingreso esperado es de 618,96 dólares; es decir, un 21,5% más que el segundo perfil, que comparte el mismo número de años de experiencia, pero únicamente alcanzó educación media.

La experiencia potencial exhibe el patrón cóncavo descrito más arriba: su término lineal ($\beta = 0{,}0290$) es positivo, cada año adicional de experiencia incrementa el ingreso en 2,9% en promedio, mientras que su término cuadrático ($\beta = -0{,}0005$) es negativo, confirmando que ese incremento se atenúa a medida que la experiencia se acumula, hasta un punto en el que, en teoría, los años adicionales de trayectoria laboral dejan de traducirse en mayores ingresos. Puntualmente, en la primera mitad de la trayectoria laboral, el incremento es marcado: un trabajador con cinco años de experiencia tiene un ingreso esperado de 457,06 dólares, mientras que uno con quince años de experiencia (diez años más, pero con la misma escolaridad) alcanza 554,33 dólares, lo que representa un incremento del 21,3%. En la segunda mitad de la trayectoria, en cambio, el mismo salto de diez años ya no produce ganancia alguna: un trabajador con veinticinco años de experiencia tiene un ingreso esperado de 609,67 dólares, y uno con treinta y cinco años de experiencia predice apenas 608,08 dólares, una variación de -0,3%, prácticamente nula.

Esta desaceleración refleja que el modelo alcanza su punto máximo teórico a los 29,7 años de experiencia potencial, y que el 43,3% de los trabajadores de la muestra objetivo se ubica en ese tramo de la trayectoria laboral o más allá de él, de modo que la porción descendente de la curva describe a una parte sustancial, no marginal, de la fuerza laboral ecuatoriana.

Siguiendo con la interpretación de los coeficientes del Modelo Clásico de Mincer, el coeficiente de área rural, finalmente, es negativo y significativo ($\beta = -0{,}0939$): un trabajador rural percibe, en promedio, un 9,0% menos que uno urbano con la misma escolaridad, la misma experiencia y la misma condición de formalidad. Esta magnitud es menor que la de la informalidad (55,0%), pero, de hecho, supera a los dos coeficientes de capital humano: es más del doble del efecto de un año adicional de escolaridad (4,0%) y

más de tres veces el de un año adicional de experiencia (2,9%). El área de residencia, en otras palabras, pesa más sobre el ingreso laboral que cualquiera de las dos variables centrales de la teoría del capital humano consideradas individualmente en cada año, aunque queda muy por debajo del efecto acumulado de una vida laboral completa de escolaridad o de experiencia.

De manera conjunta, el modelo en su versión clásica explica el 37,7% de la variabilidad del logaritmo del ingreso laboral en la muestra de entrenamiento ($R^2 = 0,377$) y prácticamente la misma proporción en la muestra de prueba ($R^2 = 0,381$), lo que descarta cualquier problema de sobreajuste y confirma que el modelo generaliza de forma adecuada fuera de la muestra con la que fue estimado. Respecto al dimensionamiento de este R^2, conviene detenerse en qué significa exactamente este valor, porque una lectura apresurada podría interpretarlo como un desempeño modesto cuando, en realidad, se ubica dentro e incluso por encima del rango que la literatura empírica reporta de forma consistente para este tipo de ecuación. Estudios que aplican la especificación de Mincer en países como México, Perú, Brasil y el propio Ecuador reportan valores de R^2 que típicamente oscilan entre 0,3 y 0,6, dependiendo de la especificación y de la calidad de las variables disponibles en la encuesta de hogares utilizada; López Lapo y Sarmiento Castillo (2019), en una aplicación directa de la ecuación de Mincer sobre la ENEMDU ecuatoriana, se ubican precisamente en ese rango. La razón de este techo relativamente bajo no es una limitación del modelo, sino un reflejo honesto de cómo se determina el ingreso laboral en la práctica: buena parte de su variación depende de atributos que ninguna encuesta de hogares releva (la habilidad individual, la calidad efectiva de la educación recibida, las redes de contactos, la suerte en la búsqueda de empleo, etc.), y Card (1999), en su revisión ya clásica de la literatura sobre retornos laborales, documenta que esta brecha entre lo observable y lo determinante del ingreso es una regularidad presente incluso en los países con los sistemas estadísticos más completos. Un R^2 cercano a 0,6 se considera, en este tipo de ejercicio, un resultado alto, precisamente porque implica explicar más de la mitad de una variabilidad que, en gran medida, escapa a lo que cualquier encuesta puede medir.

Esta observación no es un simple matiz técnico: es la razón por la que la ecuación de Mincer sigue siendo la herramienta preferida para este tipo de análisis pese a que nunca aspira a predecir el ingreso de cada trabajador en su totalidad. El ingreso laboral medido en encuestas de hogares latinoamericanas es, además, intrínsecamente volátil de una persona a otra incluso dentro de la misma categoría de trabajo, por lo que un modelo que pretendiera explicar toda esa variación tendría, en la práctica, que ajustarse por separado a cada tramo de la distribución del ingreso. Lo que la ecuación de Mincer ofrece en su lugar, y en lo que reside su ventaja comparativa frente a alternativas más complejas, es una manera transparente de atribuir la contribución relativa de cada característica observable al ingreso esperado, sin necesidad de explicar toda su variabilidad para que dicha atribución sea válida. Bajo esta consideración, el ejercicio de esta sección cumple su cometido: identifica con precisión estadística el signo, la magnitud y la significancia de los determinantes que sí es posible observar, a la vez que garantiza resultados estables, sin sobreajuste y aplicables a nuevas muestras.

Queda, sin embargo, una limitación específica de esta primera especificación, distinta de su R^2: el coeficiente de informalidad estimado aquí no distingue entre lo que es penalización propia de la informalidad y lo que responde, en realidad, a que los trabajadores informales del Ecuador se concentran de forma sistemática en establecimientos más pequeños, en ciertas ramas de actividad y en determinadas categorías ocupacionales, dimensiones que esta ecuación clásica todavía no controla. Aislar esa diferencia exige incorporar esos controles estructurales, que es precisamente el objetivo de la especificación extendida que se desarrolla a continuación.

4.2.2. La ecuación extendida: hacia una lectura estructural del ingreso laboral

La subsección anterior dejó planteada una pregunta concreta: de ese 55,0% de penalización bruta que arroja la ecuación clásica, ¿cuánto responde efectivamente a la informalidad y cuánto responde a que los trabajadores informales del Ecuador tienden a concentrarse en los segmentos menos favorecidos de otras dimensiones, como por ejemplo respecto al tamaño del establecimiento en el que trabajan, al sector económico al que pertenecen, a la categoría ocupacional que ejercen, e incluso a la posición socioeconómica del hogar del que provienen? Esta sección responde a esa pregunta ampliando la ecuación de Mincer Clásica y complementándola con la que se denomina una ecuación de Mincer Extendida, que incluye ese conjunto de controles estructurales, bajo una lógica sencilla: Si, una vez incorporados esos controles, el coeficiente de informalidad se reduce pero se mantiene negativo y significativo, ese remanente puede interpretarse como una aproximación precisa de la penalización neta que la condición de informalidad impone, más allá de las circunstancias estructurales en las que cada trabajador se desempeña.

La versión extendida de Mincer que se ha diseñado incorpora seis bloques de variables nuevas, cada uno con una motivación propia. El sexo del trabajador, abreviado como mujer (uno si es mujer, cero si es hombre), permite verificar si la brecha de género que la caracterización de la sección 4.1 documentó a nivel descriptivo persiste una vez controladas las demás covariables. Esta es una pregunta relevante en sí misma, dado que, como documentan Blau y Kahn (2017) en su revisión global de la brecha salarial de género, buena parte de esa brecha persiste, no desaparece ni se reduce significativamente en la mayoría de los países, con la incorporación de controles por educación y experiencia. El estado civil, construido a partir de cinco categorías (casado/a, divorciado/a, separado/a, unión libre y viudo/a, todas comparadas con soltero/a, que actúa como referencia), captura diferencias de ingreso asociadas a la estructura del hogar. El tamaño del establecimiento en el que trabaja cada persona, recodificado en seis categorías que van desde microempresas de diez a diecinueve empleados hasta establecimientos de setenta y cinco o más, todas comparadas con la microempresa de uno a nueve empleados como referencia, permite aislar si el ingreso depende de la escala de la unidad productiva y no solo de las características del propio trabajador. El grupo ocupacional (patrono, cuenta propia, jornalero, trabajo no remunerado y servicio doméstico, comparados con el empleado privado como categoría de referencia) captura el tipo de inserción laboral específica de cada persona. La rama de actividad económica, con hasta veinte sectores comparados con la agricultura y la pesca, siendo la rama de referencia y la de mayor incidencia de informalidad del país, según se documentó en la sección de dimensionamiento y caracterización de la informalidad, controla que la penalización atribuible a la informalidad no se sobrestima por el sector productivo en el que cada trabajador se desempeña. Y el quintil de ingreso per cápita del hogar, comparado con el quintil más pobre, incorpora la posición socioeconómica del hogar de origen como determinante estructural adicional del ingreso individual. Con estas consideraciones, la forma funcional de la ecuación de Mincer extendida que se empleó en el presente estudio es la siguiente:

$$log_ingrl = \beta_0 + \beta_1 \cdot es_informal + \beta_2 \cdot year_sh + \beta_3 \cdot exp_sh + \beta_4 \cdot exp_sh^2 + \beta_5 \cdot mujer + \beta_6 \cdot rural + \Sigma \gamma_k \cdot est_civil_k + \Sigma \delta_j \cdot tamaño_j + \Sigma \theta_m \cdot ocupación_m + \Sigma \rho_r \cdot rama_r + \Sigma \varphi_\varphi \cdot quintil_\varphi + \varepsilon$$

Donde: cada término con subíndice representa la suma de los coeficientes asociados a las variables dicotómicas de esa categoría. Por ejemplo, $\Sigma\delta_j \cdot tamaño_j$ agrupa los seis coeficientes correspondientes a las seis categorías de tamaño del establecimiento distintas de la categoría de referencia.

Dentro de la estructura del modelo empleado (Mincer Extendido), la inclusión del quintil de ingreso del hogar constituye una decisión metodológica que conviene explicitar antes de presentar los resultados: al

competir con la escolaridad por explicar una porción común de la varianza del ingreso, el quintil de ingreso del hogar está, después de todo, parcialmente determinado por la educación de sus miembros. En consecuencia, la escolaridad puede perder significancia estadística en esta versión del modelo. Se trata de un costo asumido deliberadamente, pues el objetivo de este estudio no es aislar el retorno a la educación en sí mismo, para lo cual ya existe una vasta literatura propia, sino aprovechar la ecuación de Mincer como plataforma válida para aislar el efecto de la informalidad sobre el ingreso (este es el fin último), controlando por el mayor número posible de factores estructurales relevantes.

Teniendo esta perspectiva clara y las consideraciones antes mencionadas, presentes; el modelo se estima de forma equivalente que, en la sección anterior, por mínimos cuadrados ordinarios (OLS) sobre el 70% de entrenamiento de la muestra objetivo. En este caso, 77.785 observaciones, una cifra menor que las 96.311 de la ecuación clásica, porque la especificación extendida exige información completa sobre el estado civil, el tamaño del establecimiento y el grupo ocupacional, no disponible en todos los casos. El Cuadro 9 presenta la totalidad de los coeficientes estimados en el mismo formato que el Cuadro 8.

Cuadro 9. Ecuación de Mincer extendida (conjunto de entrenamiento).

R^2 = 0,492 | R^2 ajustado = 0,492 | Estadístico F = 1641. | N (entrenamiento) = 77,785

Variable	Coef.	Error est.	t	P>|t|	IC 95%
Constante	4,9917	0,016	309,237	<0,001	[4,960, 5,023]
Informalidad	-0,3670	0,006	-56,887	<0,001	[-0,380, -0,354]
Escolaridad (year_sh)	0,0003	0,001	0,442	0,658	[-0,001, 0,002]
Experiencia potencial (exp_sh)	0,0213	0,000	44,906	<0,001	[0,020, 0,022]
Experiencia potencial² (exp_sh²)	-0,0004	6,92e-06	-63,694	<0,001	[-0,000, -0,000]
Mujer	-0,3679	0,005	-68,362	<0,001	[-0,378, -0,357]
Área rural	0,0259	0,006	4,119	<0,001	[0,014, 0,038]
Estado civil: casado/a	0,1925	0,007	27,997	<0,001	[0,179, 0,206]
Estado civil: divorciado/a	0,1063	0,012	8,873	<0,001	[0,083, 0,130]
Estado civil: separado/a	0,1357	0,009	14,298	<0,001	[0,117, 0,154]
Estado civil: unión libre	0,2075	0,007	29,952	<0,001	[0,194, 0,221]
Estado civil: viudo/a	0,1033	0,013	7,762	<0,001	[0,077, 0,129]
Tamaño: grande (≥75 empl.)	0,3584	0,037	9,603	<0,001	[0,285, 0,431]
Tamaño: mediana (50-60 empl.)	0,2925	0,019	15,678	<0,001	[0,256, 0,329]
Tamaño: mediana (61-75 empl.)	0,3479	0,049	7,144	<0,001	[0,252, 0,443]
Tamaño: microempresa (10-19 empl.)	0,1731	0,011	15,155	<0,001	[0,151, 0,196]
Tamaño: pequeña (20-35 empl.)	0,2452	0,012	19,841	<0,001	[0,221, 0,269]
Tamaño: pequeña (36-49 empl.)	0,3051	0,024	12,505	<0,001	[0,257, 0,353]
Ocupación: cuenta propia	-0,1245	0,007	-16,998	<0,001	[-0,139, -0,110]

Ocupación: jornalero	0,1238	0,010	12,157	<0,001	[0,104, 0,144]
Ocupación: patrono	0,3727	0,014	27,128	<0,001	[0,346, 0,400]
Ocupación: servicio doméstico	-0,4890	0,453	-1,080	0,280	[-1,376, 0,398]
Ocupación: trabajo no remunerado	-0,7095	0,032	-22,102	<0,001	[-0,772, -0,647]
Rama: Administración pública	0,6171	0,057	10,794	<0,001	[0,505, 0,729]
Rama: Agua y alcantarillado	0,1452	0,062	2,339	0,019	[0,024, 0,267]
Rama: Alojamiento y comida	0,3615	0,011	32,445	<0,001	[0,340, 0,383]
Rama: Artes y recreación	0,1398	0,025	5,536	<0,001	[0,090, 0,189]
Rama: Comercio y vehículos	0,2772	0,009	30,679	<0,001	[0,259, 0,295]
Rama: Construcción	0,2986	0,010	28,665	<0,001	[0,278, 0,319]
Rama: Electricidad y gas	-0,2828	0,101	-2,788	0,005	[-0,482, -0,084]
Rama: Enseñanza	0,2497	0,018	13,516	<0,001	[0,213, 0,286]
Rama: Financieras y seguros	0,5034	0,040	12,736	<0,001	[0,426, 0,581]
Rama: Hogares privados	0,7139	0,453	1,577	0,115	[-0,173, 1,601]
Rama: Información y comunicación	0,3294	0,026	12,503	<0,001	[0,278, 0,381]
Rama: Inmobiliarias	0,3743	0,034	10,912	<0,001	[0,307, 0,441]
Rama: Manufacturas	0,1924	0,010	18,382	<0,001	[0,172, 0,213]
Rama: Minas y canteras	0,3417	0,031	11,096	<0,001	[0,281, 0,402]
Rama: Organizaciones extraterritoriales	0,4515	0,320	1,410	0,158	[-0,176, 1,079]
Rama: Otros servicios	0,1498	0,014	10,679	<0,001	[0,122, 0,177]
Rama: Profesionales y técnicas	0,4032	0,016	25,965	<0,001	[0,373, 0,434]
Rama: Salud y sociales	0,3719	0,018	21,187	<0,001	[0,338, 0,406]
Rama: Servicios administrativos	0,1455	0,014	10,081	<0,001	[0,117, 0,174]
Rama: Transporte	0,3667	0,011	32,673	<0,001	[0,345, 0,389]
Quintil de ingreso del hogar 2	0,3978	0,008	50,014	<0,001	[0,382, 0,413]
Quintil de ingreso del hogar 3	0,5898	0,008	74,698	<0,001	[0,574, 0,605]
Quintil de ingreso del hogar 4	0,7768	0,008	95,503	<0,001	[0,761, 0,793]
Quintil de ingreso del hogar 5	1,0448	0,009	116,534	<0,001	[1,027, 1,062]

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.2.

El modelo explica el 49,3% de la variabilidad del ingreso en entrenamiento y el 49,1% en prueba (R^2 = 0,493 y 0,491, respectivamente), una mejora considerable frente al 37,7% de la ecuación clásica y un

valor que, siguiendo la referencia regional discutida en la subsección anterior, se ubica en el extremo superior del rango esperado para este tipo de aplicaciones de Mincer en encuestas de hogares. Con este antecedente, la interpretación de los coeficientes sigue el mismo orden en que las variables fueron presentadas: primero la informalidad, luego el capital humano propio, después las características demográficas y del hogar, y finalmente los tres bloques estructurales (tamaño del establecimiento, grupo ocupacional y rama de actividad), que cierran con el quintil de ingreso del hogar.

Según muestra el planeamiento de Mincer Extendido, el coeficiente de informalidad se mantiene negativo y altamente significativo (β = -0,3670), equivalente a una penalización de 30,7% sobre el ingreso esperado. Su magnitud es considerablemente menor que el 55,0% de la ecuación clásica, y esa reducción es, en sí misma, un hallazgo relevante, pues indica que buena parte de la brecha bruta observada entre formales e informales responde, efectivamente, a que estos últimos se concentran en establecimientos más pequeños, en ocupaciones de menor remuneración y en hogares de menor posición socioeconómica, y no a la informalidad en sí misma. El 30,7% que resta, sin embargo, no desaparece: persiste incluso al comparar trabajadores con el mismo nivel educativo, la misma experiencia, el mismo tamaño de establecimiento y la misma inserción ocupacional, lo que sugiere que hay algo en la condición de informalidad —la ausencia de un contrato formal, de afiliación a la seguridad social y de mecanismos de negociación salarial— que el mercado laboral ecuatoriano penaliza de manera directa.

Por su parte, los coeficientes de capital humano propio conservan, con una salvedad, el comportamiento ya observado en la ecuación clásica. La experiencia potencial mantiene su patrón cóncavo (β = 0,0213 en el término lineal y -0,0004 en el cuadrático), con una magnitud ligeramente menor que antes, pero en la misma dirección. La escolaridad, en cambio, pierde su significancia estadística (β = 0,0003, p = 0,658). Se sospecha de la existencia de colinealidad entre esta variable y el quintil de ingreso del hogar; sin embargo, esta sospecha se aborda específicamente en el Anexo V, en el que se ejecuta un análisis de Factor de Inflación de la Varianza (VIF) para confirmar o descartar este indicio. El resultado descarta colinealidad severa entre ambas variables (VIF = 2,52 para escolaridad y entre 1,89 y 2,42 para las categorías de quintil, todos dentro del rango 1-5 considerado aceptable, muy por debajo del umbral de 10 que señalaría un problema estadístico grave). La pérdida de significancia de la escolaridad responde, en consecuencia, no a una inestabilidad estadística de los coeficientes por colinealidad severa, sino a que ambas variables comparten una porción sustantiva de la variación del ingreso que explican (el quintil de ingreso del hogar, al ser en parte una función de la escolaridad de sus miembros, capta una porción de esa misma información), lo que reduce el aporte marginal e independiente de la escolaridad una vez que el quintil ya está presente en el modelo. Quien desee evaluar el retorno a la educación de forma aislada puede remitirse a la ecuación de Mincer clásica presentada en la subsección anterior, donde la escolaridad sí conserva su significancia estadística (β = 0,0388, equivalente a 4,0% por año), precisamente porque esa especificación no incluye el quintil de ingreso del hogar con el que esta variable comparte información en la ecuación extendida.

Dentro del segmento de características demográficas, el coeficiente de sexo (mujer) es negativo y muy significativo (β = -0,3679, equivalente a -30,8%): la brecha de género no solo persiste bajo este control exhaustivo de características estructurales, sino que su magnitud es casi idéntica a la de la propia informalidad, lo que sugiere que ambas condiciones operan como ejes de estratificación salarial de peso comparable en el mercado laboral ecuatoriano. El coeficiente de área rural, que en la ecuación clásica era negativo, cambia de signo en esta especificación (β = 0,0259, equivalente a +2,6%): la desventaja que antes se atribuía al área de residencia se redistribuye, en la ecuación extendida, hacia la rama de actividad y, sobre todo, hacia el quintil de ingreso del hogar, que capturan de forma más directa la fuente real de esa desventaja. Es decir, la ruralidad en sí misma no implica una penalización salarial, sino que

esta se origina en las ramas de actividad en las que el trabajador se inserta y en las condiciones de vida de los hogares, reflejadas en el ingreso familiar.

Continuando con la desagregación de Mincer Extendido, entre las categorías de estado civil, comparadas con soltero/a, las cinco resultan positivas y significativas al 1%, con las dos formas de convivencia en pareja (unión libre (β = 0,2075, +23,1%) y casado/a (β = 0,1925, +21,2%)) por delante de las categorías asociadas a una ruptura o pérdida conyugal (separado/a (+14,5%), divorciado/a (+11,2%) y viudo/a (+10,9%)). Un ejemplo concreto permite cuantificar estas diferencias en términos monetarios. Si se considera un hombre urbano, con doce años de escolaridad y diecisiete años de experiencia potencial, empleado privado en una microempresa del sector agrícola, proveniente de un hogar de clase media; si esta persona es soltera, el modelo predice un ingreso laboral esperado de 336,20 dólares. Sí, manteniendo idénticas todas las demás características, si vive en unión libre, el ingreso esperado sube a 413,95 dólares (23,1% más); si está casada, a 407,78 dólares (21,3% más). En el otro extremo, si está separada, el ingreso esperado es de 385,19 dólares (14,6% más que el de un soltero); si está divorciada, el ingreso pasa a ser de 373,99 dólares (11,2% más); y si es viuda, el ingreso será de 372,88 dólares (10,9% más). La brecha completa entre la categoría de mayor ingreso (unión libre) y la de menor ingreso dentro de este grupo (viudo/a) asciende a 41,07 dólares mensuales, pese a tratarse, en los seis casos, de la misma persona con la misma escolaridad, experiencia, ocupación y posición socioeconómica del hogar. Este patrón es consistente con la interpretación habitual en la literatura de que la convivencia en pareja actúa como un incentivo indirecto para una mayor inserción laboral, más que como una relación causal directa entre el estado civil y la productividad del trabajador.

Por su parte, el tamaño del establecimiento exhibe una prima salarial que se observa con claridad incluso dentro de la microempresa: pasar de un establecimiento de uno a nueve empleados a uno de diez a diecinueve ya se asocia con un ingreso 18,9% mayor, y esa prima continúa creciendo hasta alcanzar 43,1% en los establecimientos de setenta y cinco empleados o más. La consistencia de este gradiente a través de las seis categorías sugiere que no se trata de un efecto aislado de la gran empresa, sino de una prima que aumenta de forma prácticamente continua con la escala de la unidad productiva, lo cual es coherente con que las firmas más grandes suelen operar con mayores márgenes de productividad por trabajador, mejor acceso al capital y mayor formalización de sus relaciones laborales internas.

Por su parte, el grupo ocupacional concentra los coeficientes de mayor magnitud en ambas direcciones del modelo. En este caso, como se mencionó anteriormente, la categoría de referencia es la del empleado asalariado privado. Entonces, la categoría Patrono presenta la prima más alta (β = 0,3727, +45,2%), lo cual es coherente con que el excedente de una unidad productiva beneficie a quien la dirige, independientemente de su condición de formalidad. Así, un patrono con capacidades y un contexto socioeconómico similares gana un 45,2% más que un empleado privado. Un jornalero, en cambio, muestra un coeficiente positivo pero mucho más modesto (β = 0,1238, +13,2%). El coeficiente asociado al cuentapropismo es negativo (β = -0,1245; -11,7%). Sin embargo, ello no implica que el ingreso total del cuentapropista sea necesariamente menor que el de un empleado privado, sino que este está afectado por la estructura de costos de su unidad productiva y, cuando dicha estructura es alta, el ingreso disponible se reduce. En términos agregados, el modelo muestra que el ingreso laboral de un cuentapropista con capacidades y circunstancias equivalentes a las de un empleado privado es 12% menor, en promedio. El coeficiente más negativo de todo el modelo corresponde al trabajo no remunerado (β = -0,7095, -50,8%), mientras que el servicio doméstico, pese a un coeficiente de magnitud considerable (-38,7%), no alcanza significancia estadística (p = 0,280), lo que refleja el reducido número de observaciones de esta categoría en la muestra.

En esta misma línea, la rama de actividad económica, comparada con la agricultura y la pesca, confirma que prácticamente cualquier otro sector paga mejor que el agrícola: la administración pública encabeza la lista (+85,3%), seguida de las financieras y de seguros (+65,4%) y de las carreras profesionales y técnicas (+49,7%). Construcción merece una mención aparte: pese a exponer a sus trabajadores a uno de los riesgos de informalidad más altos del país, según se documentará en la sección de caracterización y dimensionamiento de la informalidad, su ingreso promedio (+29,9%) se ubica por encima del de sectores considerablemente menos riesgosos, lo que evidencia que la estructura de riesgos y la de ingresos de una rama no siempre se mueven en la misma dirección.

El quintil de procedencia del hogar cierra la interpretación con el coeficiente de mayor magnitud de todo el modelo: pertenecer a un hogar del quintil más rico se asocia con un ingreso individual 184,3% superior al del quintil más pobre ($\beta = 1,0448$), con un gradiente estrictamente monótono en las categorías intermedias: -48,9% en el quintil 2, 80,4% en el quintil 3 y 117,4% en el quintil 4. Esta magnitud, muy por encima de la de cualquier otra variable del modelo, confirma lo que la sección 4.1 ya insinuaba desde una óptica descriptiva: la posición socioeconómica del hogar de origen pesa, en la determinación del ingreso laboral de un individuo, tanto o más que cualquier característica propia de su puesto de trabajo. Para ilustrarlo, considérese, una última vez, al mismo hombre urbano, soltero, formal, con doce años de escolaridad, diecisiete años de experiencia potencial, empleado privado en una microempresa agrícola: si proviene de un hogar del quintil de ingreso más pobre, el modelo predice un ingreso esperado de apenas 185,95 dólares; si proviniera de un hogar del quintil más rico, manteniendo idéntico el resto de su perfil, el ingreso esperado sube a 530,49 dólares, casi 345 dólares más al mes. Entre ambos extremos, el ascenso es progresivo y constante: 277,29 dólares en el quintil 2, 336,20 en el quintil 3 y 405,52 en el quintil 4. El salto completo entre el origen más pobre y el más rico (185,3%) supera así, por un margen amplio, tanto a la brecha de género como a la propia penalización de la informalidad estimada en esta misma ecuación, lo que confirma que ninguna decisión tomada por el propio trabajador dentro de su vida laboral pesa tanto sobre su ingreso como el hogar en el que nació.

Como consecuencia de lo expuesto, la Figura 12 ilustra el ajuste global alcanzado por esta especificación de Mincer al comparar, sobre el conjunto de prueba, la distribución del logaritmo del ingreso entre los trabajadores informales (observada y predicha por el modelo) frente a la distribución observada entre los trabajadores formales. La curva de valores predichos para los ocupados informales se ajusta de manera razonablemente cercana a la observada, y ambas se ubican de forma consistente a la izquierda de la distribución formal, lo que constituye evidencia visual directa de la brecha y la penalización salarial de este segmento poblacional.

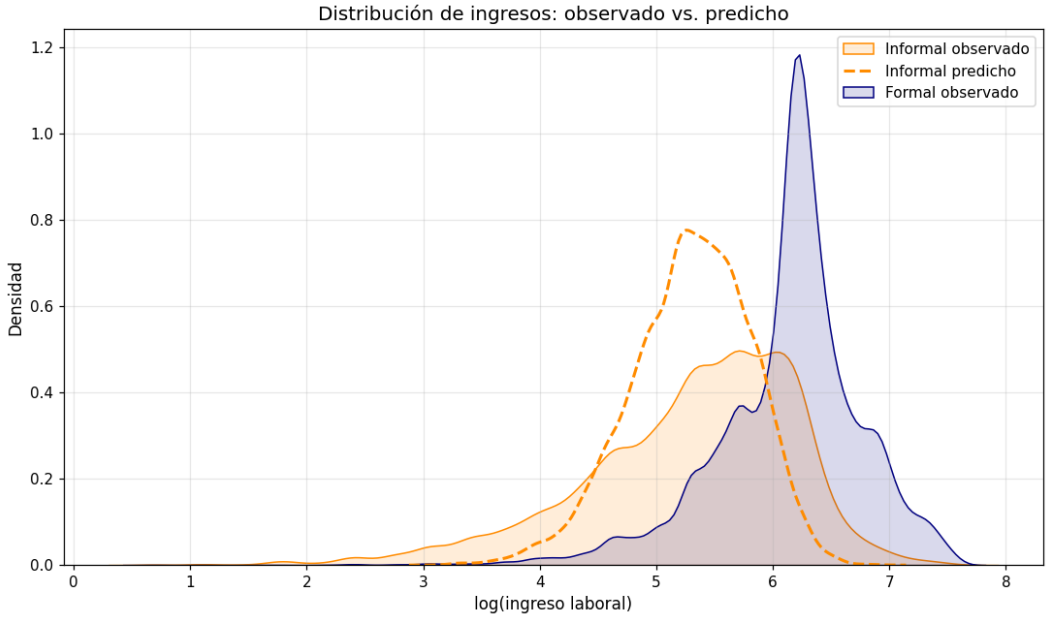


Figura 12. Distribución del logaritmo del ingreso laboral, observado y predicho, formal vs. informal. Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.2.

En este punto es pertinente detenerse en lo que este conjunto de resultados dice, en su totalidad, sobre el rostro del sistema laboral ecuatoriano. El ingreso de un trabajador no depende únicamente de su esfuerzo formativo ni de su condición de formalidad, sino de una combinación de circunstancias que exceden su control inmediato — el sector en el que su economía local le permite insertarse, la escala de la firma que lo contrata, el quintil de ingreso del hogar en el que nació— y de decisiones más próximas a su propia trayectoria, como la categoría ocupacional que asume: ser patrono, trabajador por cuenta propia o empleado. La informalidad es una de estas variables, no la única, y presentarla de manera aislada —como si el resto de los factores no operara simultáneamente— distorsionaría el diagnóstico. Pero es, entre todas ellas, una de las de mayor magnitud individual, comparable a la brecha de género y superior al efecto marginal de un año adicional de escolaridad. La evidencia que se acumula a lo largo de esta sección respalda la idea de que, exista o no la informalidad como alternativa atractiva de corto plazo —para maximizar el margen de un pequeño negocio, para evitar el desempleo abierto, para ingresar con rapidez a un mercado laboral que de otro modo exige credenciales o contactos—, ese atractivo convive con un costo salarial cierto y sustancial para quien la ejerce.

4.2.3. Validación desde un enfoque alternativo: la descomposición de Oaxaca-Blinder

La ecuación extendida de la subsección anterior descansa sobre un supuesto que hasta ahora no se ha puesto a prueba: que el retorno a cada característica (un año adicional de escolaridad, un año adicional de experiencia, trabajar en una empresa más grande), es idéntico para trabajadores formales e informales, y que la única diferencia entre ambos grupos es el intercepto que captura la dummy de informalidad. Ese supuesto es razonable como punto de partida, pero no es necesariamente cierto: es perfectamente posible que un año adicional de escolaridad valga más en el sector formal que en el informal, o que el tamaño del establecimiento modifique el ingreso de manera distinta según la condición

de formalidad del trabajador. Esta subsección introduce una técnica capaz de prescindir totalmente de ese supuesto. Se denomina Modelo de Oaxaca (1973) y Blinder (1973), y su utilidad radica en que, desde un ángulo metodológicamente independiente, permite determinar si la penalización del 30,7% estimada en la subsección anterior es un hallazgo robusto que puede ser confirmado y validado desde una lógica de construcción distinta.

La lógica de la técnica es la siguiente: en lugar de estimar una única ecuación con una dummy de informalidad, se estima una ecuación de Mincer extendida por separado para cada grupo, una para los trabajadores formales, otra para los trabajadores informales, ambas con exactamente las mismas covariables de la subsección anterior, salvo la propia dummy de informalidad, que en este caso deja de ser necesaria porque se define directamente a qué submuestra pertenece cada observación. Con esas dos ecuaciones estimadas, la brecha salarial promedio entre ambos grupos (la diferencia entre el ingreso promedio de los formales y el de los informales, en logaritmos) se descompone en dos componentes.

El primero, llamado efecto dotaciones, mide el efecto del déficit de condiciones o características que, en promedio, presenta un trabajador informal frente a uno formal: menor escolaridad, establecimientos más pequeños, ocupaciones peor remuneradas. Es decir, captura la brecha que existiría incluso si el mercado pagara exactamente lo mismo a ambos grupos por cada característica que presenten; como si, en una entrevista de trabajo, cada acción afirmativa y cada requisito básico de los candidatos recibieran la misma puntuación. Pese a ello, los trabajadores informales, al presentar un déficit de características; terminarían recibiendo menos puntuación, al menos en teoría y haciendo una generalización ilustrativa. Para calcularlo, se usa la misma tabla de puntuación (del sector formal) para calificar a ambos candidatos. Al trabajador informal representativo se le puntúa según sus propias características (nueve años de escolaridad, microempresa pequeña, cuenta propia), pero utilizando la tabla de puntuación formal. Como resultado, obtiene un puntaje equivalente a 212,19 dólares. Al trabajador formal representativo se le puntúa según sus propias características (13 años de escolaridad, empresa algo más grande, empleado privado), con la misma tabla: obtiene 490,14 dólares. La diferencia entre ambos resultados (277,95 dólares, un 131,0%) no puede deberse a que los evaluaron con reglas distintas, porque se usó la misma tabla de puntuación para ambos; solo puede deberse a que uno de los dos candidatos llegó con mejores credenciales. Esa diferencia es precisamente el efecto de las dotaciones.

Un ejemplo con perfiles representativos de la ENEMDU permite ver la brecha completa de principio a fin. Considérese, por un lado, a un trabajador informal representativo: cuenta propia, nueve años de escolaridad, en una microempresa de uno a nueve empleados, y perteneciente al quintil de ingreso más bajo del hogar. Este trabajador percibe, con la estructura de retornos que efectivamente rige en el sector informal, un ingreso de 160,86 dólares. Por otro lado, un trabajador formal representativo -empleado privado, trece años de escolaridad, en una microempresa algo más grande (de diez a diecinueve empleados), en el tercer quintil de ingreso del hogar- percibe, con la estructura de retornos del sector formal, 490,14 dólares. La brecha entre ambos es de 329,28 dólares, lo que equivale a que el trabajador formal gana un 204,7% más que el trabajador informal.

El segundo componente, llamado efecto de coeficientes, aísla lo opuesto: cuánto de la brecha se debe a que el mercado paga distinto por las mismas características, y no a que esas características sean distintas. Para calcularlo, se toma al mismo trabajador informal representativo, con sus propias características fijas, y se compara su ingreso real (calculado con la estructura de retornos informal) con el ingreso que tendría esa misma persona, con esas mismas características, si el mercado le pagara con la estructura de retornos del sector formal. Como en ambos casos se trata de la misma persona con las mismas dotaciones, cualquier diferencia entre los dos resultados no puede deberse a diferencias de

características: solo puede deberse a que el mercado retribuye esas características de manera distinta según la condición de formalidad.

Retomando al mismo trabajador informal representativo del ejemplo anterior (cuenta propia, nueve años de escolaridad, microempresa de uno a nueve empleados, quintil de ingreso más bajo): con su propia estructura de retornos informales, su ingreso real es de 160,86 dólares. Sí, exactamente esas mismas características (ni una más ni una menos) se pagarían con la estructura de retornos formal, el ingreso subiría a 212,19 dólares. Esa diferencia de 51,33 dólares, equivalente al 31,9%, es el efecto de los coeficientes: la penalización que ese trabajador enfrenta por el solo hecho de ser informal, ya que ninguna de sus características cambió entre un cálculo y otro, solo la estructura de retornos con la que el mercado las remunera.

Con estas consideraciones de partida, las regresiones formal e informal cumplen dos propósitos distintos, y cada una exige un tratamiento distinto de los datos. El primero es validar que ambas ecuaciones generalizan bien, para lo que cada una se estima sobre el 70% de entrenamiento de su respectiva submuestra; esos son los resultados que reportan los Cuadros 9 y 10. El segundo es calcular la descomposición de Oaxaca-Blinder en sí misma, que no es un ejercicio de predicción a evaluar fuera de muestra, sino una estimación puntual de la brecha promedio entre dos poblaciones y de cómo esta se reparte entre dotaciones y coeficientes. Para este segundo propósito, ambas ecuaciones se reestiman sobre la totalidad de cada submuestra (es decir, el 100% de los registros).

Como este segundo ejercicio no es una predicción que deba ponerse a prueba fuera de muestra, sino una estimación puntual sobre la totalidad de los datos disponibles, reservar un 30% para prueba no aportaría ninguna garantía adicional y solo restaría precisión al cálculo; por esa razón, y siguiendo la práctica estándar de la literatura sobre esta técnica, es esta segunda versión, estimada sobre cada submuestra completa, la que alimenta la descomposición de Oaxaca-Blinder que se presenta en el Cuadro 12.

Cuadro 10. Regresión de Mincer extendida en el sector formal (conjunto de entrenamiento).

R^2 = 0,384 | R^2 ajustado = 0,384 | Estadístico F = 540,7 | N (entrenamiento) = 38,130

Variable	Coef.	Error est.	t	P>|t|	IC 95%
Constante	5,1293	0,021	248,686	<0,001	[5,089, 5,170]
Escolaridad (year_sh)	0,0060	0,001	6,937	<0,001	[0,004, 0,008]
Experiencia potencial (exp_sh)	0,0194	0,001	33,694	<0,001	[0,018, 0,020]
Experiencia potencial² (exp_sh²)	-0,0004	9,67e-06	-39,834	<0,001	[-0,000, -0,000]
Mujer	-0,1922	0,006	-33,117	<0,001	[-0,204, -0,181]
Área rural	-0,0064	0,007	-0,864	0,387	[-0,021, 0,008]
Estado civil: casado/a	0,1718	0,008	22,682	<0,001	[0,157, 0,187]
Estado civil: divorciado/a	0,0585	0,013	4,606	<0,001	[0,034, 0,083]
Estado civil: separado/a	0,1070	0,011	9,450	<0,001	[0,085, 0,129]
Estado civil: unión libre	0,1602	0,008	20,631	<0,001	[0,145, 0,175]
Estado civil: viudo/a	0,0796	0,018	4,313	<0,001	[0,043, 0,116]

Tamaño: grande (≥75 empl.)	0,3537	0,029	12,178	<0,001	[0,297, 0,411]
Tamaño: mediana (50-60 empl.)	0,3061	0,015	20,946	<0,001	[0,277, 0,335]
Tamaño: mediana (61-75 empl.)	0,3310	0,038	8,786	<0,001	[0,257, 0,405]
Tamaño: microempresa (10-19 empl.)	0,1917	0,010	19,953	<0,001	[0,173, 0,211]
Tamaño: pequeña (20-35 empl.)	0,2705	0,010	27,111	<0,001	[0,251, 0,290]
Tamaño: pequeña (36-49 empl.)	0,3104	0,019	15,955	<0,001	[0,272, 0,349]
Ocupación: cuenta propia	-0,0247	0,007	-3,386	0,001	[-0,039, -0,010]
Ocupación: jornalero	-0,2370	0,015	-15,893	<0,001	[-0,266, -0,208]
Ocupación: patrono	0,3386	0,012	27,596	<0,001	[0,315, 0,363]
Ocupación: servicio doméstico	-0,0468	0,008	-5,828	<0,001	[-0,063, -0,031]
Ocupación: trabajo no remunerado	-1,2657	0,054	-23,578	<0,001	[-1,371, -1,161]
Rama: Administración pública	0,2366	0,046	5,147	<0,001	[0,147, 0,327]
Rama: Agua y alcantarillado	0,0621	0,056	1,112	0,266	[-0,047, 0,172]
Rama: Alojamiento y comida	-0,0115	0,015	-0,785	0,433	[-0,040, 0,017]
Rama: Artes y recreación	-0,1290	0,027	-4,811	<0,001	[-0,182, -0,076]
Rama: Comercio y vehículos	-0,0129	0,013	-0,995	0,320	[-0,038, 0,013]
Rama: Construcción	0,0881	0,016	5,395	<0,001	[0,056, 0,120]
Rama: Electricidad y gas	0,0996	0,125	0,796	0,426	[-0,146, 0,345]
Rama: Enseñanza	-0,0441	0,019	-2,345	0,019	[-0,081, -0,007]
Rama: Financieras y seguros	0,1830	0,034	5,386	<0,001	[0,116, 0,250]
Rama: Hogares privados	-0,0468	0,008	-5,828	<0,001	[-0,063, -0,031]
Rama: Información y comunicación	0,0494	0,025	1,983	0,047	[0,001, 0,098]
Rama: Inmobiliarias	0,1012	0,031	3,252	0,001	[0,040, 0,162]
Rama: Manufacturas	-0,0543	0,014	-3,934	<0,001	[-0,081, -0,027]
Rama: Minas y canteras	0,1387	0,028	4,931	<0,001	[0,084, 0,194]
Rama: Organizaciones extraterritoriales	0,1363	0,223	0,611	0,541	[-0,301, 0,574]
Rama: Otros servicios	-0,0715	0,018	-3,876	<0,001	[-0,108, -0,035]
Rama: Profesionales y técnicas	0,0726	0,017	4,354	<0,001	[0,040, 0,105]
Rama: Salud y sociales	0,0651	0,018	3,548	<0,001	[0,029, 0,101]
Rama: Servicios administrativos	-0,0846	0,017	-4,916	<0,001	[-0,118, -0,051]
Rama: Transporte	0,0608	0,015	4,033	<0,001	[0,031, 0,090]

Quintil de ingreso del hogar 2	0,4224	0,013	33,262	<0,001	[0,397, 0,447]
Quintil de ingreso del hogar 3	0,5944	0,012	49,204	<0,001	[0,571, 0,618]
Quintil de ingreso del hogar 4	0,7588	0,012	63,627	<0,001	[0,735, 0,782]
Quintil de ingreso del hogar 5	1,0227	0,012	83,722	<0,001	[0,999, 1,047]

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.3.

Cuadro 11. Regresión de Mincer extendida, sector informal (conjunto de entrenamiento).

R^2 = 0,371 | R^2 ajustado = 0,371 | Estadístico F = 601,9 | N (entrenamiento) = 39,655

Variable	Coef.	Error est.	t	P>|t|	IC 95%
Constante	4,7654	0,026	181,991	<0,001	[4,714, 4,817]
Escolaridad (year_sh)	-0,0083	0,001	-6,701	<0,001	[-0,011, -0,006]
Experiencia potencial (exp_sh)	0,0217	0,001	28,643	<0,001	[0,020, 0,023]
Experiencia potencial² (exp_sh²)	-0,0005	1,03e-05	-45,347	<0,001	[-0,000, -0,000]
Mujer	-0,5597	0,009	-61,780	<0,001	[-0,577, -0,542]
Área rural	0,0666	0,010	6,841	<0,001	[0,048, 0,086]
Estado civil: casado/a	0,2017	0,011	17,994	<0,001	[0,180, 0,224]
Estado civil: divorciado/a	0,1467	0,021	7,004	<0,001	[0,106, 0,188]
Estado civil: separado/a	0,1613	0,015	11,011	<0,001	[0,133, 0,190]
Estado civil: unión libre	0,2283	0,011	20,575	<0,001	[0,207, 0,250]
Estado civil: viudo/a	0,1473	0,019	7,769	<0,001	[0,110, 0,184]
Tamaño: microempresa (10-19 empl.)	0,1832	0,039	4,728	<0,001	[0,107, 0,259]
Tamaño: pequeña (20-35 empl.)	0,2960	0,118	2,513	0,012	[0,065, 0,527]
Tamaño: pequeña (36-49 empl.)	-0,0466	0,736	-0,063	0,950	[-1,490, 1,397]
Ocupación: cuenta propia	-0,2018	0,015	-13,132	<0,001	[-0,232, -0,172]
Ocupación: jornalero	0,1055	0,018	5,852	<0,001	[0,070, 0,141]
Ocupación: patrono	0,4912	0,034	14,277	<0,001	[0,424, 0,559]
Ocupación: trabajo no remunerado	-0,6256	0,044	-14,328	<0,001	[-0,711, -0,540]
Rama: Administración pública	0,9836	0,279	3,531	<0,001	[0,438, 1,530]
Rama: Agua y alcantarillado	-0,0114	0,197	-0,058	0,954	[-0,398, 0,375]
Rama: Alojamiento y comida	0,5571	0,018	31,588	<0,001	[0,523, 0,592]

Rama: Artes y recreación	0,3056	0,046	6,580	<0,001	[0,215, 0,397]
Rama: Comercio y vehículos	0,3631	0,014	26,784	<0,001	[0,337, 0,390]
Rama: Construcción	0,2914	0,014	20,310	<0,001	[0,263, 0,319]
Rama: Electricidad y gas	-0,3995	0,151	-2,652	0,008	[-0,695, -0,104]
Rama: Enseñanza	0,2484	0,044	5,585	<0,001	[0,161, 0,336]
Rama: Financieras y seguros	0,8427	0,135	6,240	<0,001	[0,578, 1,107]
Rama: Información y comunicación	0,3971	0,077	5,128	<0,001	[0,245, 0,549]
Rama: Inmobiliarias	0,7206	0,119	6,074	<0,001	[0,488, 0,953]
Rama: Manufacturas	0,2670	0,017	15,840	<0,001	[0,234, 0,300]
Rama: Minas y canteras	0,3681	0,074	4,956	<0,001	[0,223, 0,514]
Rama: Otros servicios	0,2222	0,021	10,637	<0,001	[0,181, 0,263]
Rama: Profesionales y técnicas	0,6058	0,038	16,089	<0,001	[0,532, 0,680]
Rama: Salud y sociales	0,4247	0,039	10,953	<0,001	[0,349, 0,501]
Rama: Servicios administrativos	0,2867	0,023	12,433	<0,001	[0,241, 0,332]
Rama: Transporte	0,4382	0,017	25,479	<0,001	[0,405, 0,472]
Quintil de ingreso del hogar 2	0,4081	0,011	38,340	<0,001	[0,387, 0,429]
Quintil de ingreso del hogar 3	0,5904	0,011	53,685	<0,001	[0,569, 0,612]
Quintil de ingreso del hogar 4	0,7779	0,012	65,129	<0,001	[0,754, 0,801]
Quintil de ingreso del hogar 5	1,0151	0,015	67,096	<0,001	[0,985, 1,045]

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.3.

Antes de leer los resultados, se debe tener claro el mapa de lo que viene: cuatro cuadros, cada uno respondiendo una pregunta distinta. Los Cuadros 9 y 10, ya presentados, son los cimientos (las dos ecuaciones separadas, una por sector) sin los que la técnica no podría aplicarse. El Cuadro 12 es el resultado que todo el ejercicio persigue: cuánto de la brecha salarial se debe a que los formales e informales parten de circunstancias distintas, y cuánto a que el mercado los trata de manera distinta, incluso cuando esas circunstancias coinciden. El Cuadro 13, finalmente, abre esa segunda cifra por dentro: no se limita a decir cuánta brecha queda sin explicar, sino que señala, variable por variable, dónde se concentra esa diferencia de trato. Cada cuadro se construye sobre el anterior y ninguno puede reemplazar al que le sigue.

Una vez dicho esto, ambas regresiones alcanzan un ajuste similar entre sí ($R^2 = 0,384$ para formales, con 38.130 observaciones; $R^2 = 0,371$ para informales, con 39.655 observaciones), y ambos valores se ubican, nuevamente, dentro del rango esperado para este tipo de ecuación según la evidencia regional discutida en la subsección anterior. Antes de proceder a la descomposición, se debe señalar una particularidad de la regresión informal que exige cautela en su lectura: las categorías de tamaño de

establecimiento mediano y grande, y la de servicio doméstico, muestran coeficientes prácticamente nulos y errores estándar anómalos, reflejo de que, por construcción, casi no existen trabajadores informales en establecimientos de esa escala, según la metodología oficial, siendo esta una regularidad estructural, no un error de estimación, que ya se documentó al construir los perfiles de riesgo de la sección de caracterización y dimensionamiento. Esta particularidad afecta únicamente a esas categorías específicas dentro de la regresión informal, sin comprometer el resto de los coeficientes ni la descomposición agregada que se presenta a continuación.

Formalmente, la descomposición parte de la brecha promedio observada entre el logaritmo del ingreso de los formales y el de los informales, y la separa en los dos componentes ya descritos mediante la siguiente identidad:

$$\Delta = (\bar{X}_F - \bar{X}_I)'\beta_F \; + \; \bar{X}_I'(\beta_F - \beta_I)$$

$$\Delta \; = \; Efecto\ dotaciones \; + \; Efecto\ coeficientes$$

Donde: $\bar{X}_F$ y $\bar{X}_I$ son los promedios de las características observables de los sectores formal e informal (escolaridad, experiencia, tamaño del establecimiento y el resto de covariables de la ecuación extendida), y β_F y β_I son los coeficientes ya estimados en los Cuadros 9 (β_F, sector formal) y 10 (β_I, sector informal). El primer término, $(\bar{X}_F - \bar{X}_I)'\beta_F$, es el efecto de dotaciones: la diferencia entre las características promedio de ambos grupos, ponderada por los retornos del sector formal. El segundo, $\bar{X}_I'(\beta_F - \beta_I)$, es el efecto coeficientes: las características promedio de los informales, multiplicadas por la diferencia entre lo que el sector formal y el informal pagan por esas mismas características. El Cuadro 12 aplica esta identidad sobre la totalidad de cada submuestra, tanto para el 100% de trabajadores formales, como para el 100% de trabajadores informales:

Cuadro 12. Descomposición de Oaxaca-Blinder de la brecha salarial formal-informal.

Componente	**Valor (log)**	**Participación relativ**
Brecha total	0,8299	100,0%
Efecto dotaciones (explicado)	0,4066	49,0%
Efecto coeficientes (penalización pura)	0,4233	51,0%

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.3.

El resultado corrobora, desde este ángulo metodológicamente independiente, el hallazgo central de la subsección anterior: que la informalidad impone una penalización salarial real -30,7% según la ecuación extendida (que no se explica por diferencias en escolaridad, experiencia u otras características observables):

Para evidenciarlo, se debe precisar la lectura de las dos columnas del Cuadro 12, ya que, en realidad, responden a preguntas distintas. La tercera columna, denominada participación (49,0% y 51,0%), muestra la participación relativa de cada componente en la brecha total de ingresos, expresada en logaritmos. Como los valores de dotaciones (0,4066) y coeficientes (0,4233) suman exactamente una brecha total de 0,8299 las participaciones relativas se calculan dividiendo cada efecto por esta magnitud, hallando su peso relativo: 49,0% y 51,0%. Esta es una medida válida, pero debe entenderse como un reparto interno en la escala logarítmica, no como un porcentaje de ingreso ni como una magnitud monetaria.

En este contexto el hallazgo revelado mediante el Cuadro 12 merece ser interpretado con atención: El componente mayor (51,0%) que explica la brecha salarial entre forales e informales no responde a dotaciones sino a coeficientes; es decir, más de la mitad de la brecha salarial no se explica porque los informales tengan menos escolaridad, trabajen en empresas más pequeñas o se concentren en ramas menos productivas, sino porque el mercado retribuye esas mismas características de manera distinta según la condición de formalidad del trabajador. Esto tiene una implicación que debe hacerse explícita: si la informalidad fuera, ante todo, un problema de capital humano, la brecha debería explicarse casi en su totalidad por dotaciones, y sería necesario priorizar programas de educación y capacitación para cerrarla con el tiempo. Que más de la mitad de la brecha resista ese control sugiere, en cambio, que la informalidad opera como una contribución adicional onerosa que el trabajador debe pagar para mantenerse fuera del esquema de retribuciones asociado a la formalidad. Es decir, el trabajador se encuentra frente a una especie de peaje que se cobra por la mera condición de operar fuera del marco formal, y que ni la educación ni la experiencia logran evitar. Perry et al. (2007) describen exactamente este mecanismo bajo la etiqueta de informalidad de exclusión, y Robles, Ponce, Alvarado y Ortiz (2019), en una aplicación de esta misma técnica sobre la ENEMDU centrada en la brecha por sexo e informalidad, encuentran un patrón consistente: una porción sustancial de la brecha salarial asociada a la informalidad permanece sin explicación por diferencias en la dotación de capital humano.

Esta lectura tiene una implicación de política pública relevante que no debe dejarse implícita: si la barrera es principalmente de acceso, por encima de la capacidad, los instrumentos que funcionan son distintos de los que suele privilegiar el debate público. Programas de capacitación laboral o de nivelación educativa, por valiosos que sean por otras razones, abordan la mitad menos relevante/urgente del problema: actúan sobre las dotaciones, cuando el componente mayor de la brecha está en los coeficientes. Lo que este resultado demanda entonces son políticas que reduzcan el costo mismo de formalizarse, tales como: simplificación de trámites y regímenes tributarios para la microempresa, portabilidad de la seguridad social entre empleos, incentivos tributarios transitorios para la contratación formal y mecanismos que faciliten a un trabajador informal acceder a la estructura de retornos formal sin necesariamente cambiar o buscar incidir en sus características observables primero. Esto, sin considerar la inversión generacional que ello implica al moldear, potenciar o modificar características observables. Los retornos de invertir en educación por ejemplo, tardarán en surtir efecto entre 5 y 15 años. Sin embargo, el modelo nos muestra que existen soluciones menos costosas, más rápidas de implementar y con mayor impacto. Esto no quiere decir que haya que desacelerar la inversión en la potenciación de características observables o en el gasto social, sino que trabajar en ellas no garantiza que un trabajador rompa la trampa de la informalidad. El modelo más bien muestra que muchos trabajadores buenos, capaces, bien educados, y con capacidades equivalentes a las de un trabajador formal que tiene un salario adecuado, hoy se están quedando fuera de la formalidad y su salario está siendo penalizado, precisamente por las barreras de entrada monetarias, administrativas, e informativas que existen para ser parte de ella; y no porque exista un déficit de competencias.

Por otro lado, el 49,0% restante, sin embargo, no es un residuo menor: casi la mitad de la brecha sí responde a que los trabajadores informales llegan, en promedio, con un paquete de características menos favorables (menor escolaridad, postulaciones y aceptaciones principalmente en microempresas más pequeñas, o ramas de actividad de menor productividad). Este porcentaje explicado de la brecha sí es sensible a los instrumentos de política que atacan las dotaciones, tales como: el acceso a educación y capacitación de calidad, a programas de bachillerato acelerado, a programas de profesionalización virtuales, a las transferencias sociales condicionadas sujetas a mantener el equilibrio del hogar; así como el crédito productivo para que las microempresas informales crezcan de escala, y las políticas de desarrollo productivo sectorial que trasladen trabajadores hacia ramas de mayor productividad, todos

ellos con potencial de cerrar, con el tiempo, esa porción de la brecha. La lectura completa del Cuadro 12, entonces, no es que la política de capital humano sea menos útil, sino que actúa sobre menos de la mitad del problema, y que ignorar la otra mitad (de los coeficientes) condenaría a cualquier estrategia de cierre de brechas a un resultado parcial.

Respecto a la segunda columna del mismo cuadro, denominada valor log, cabe destacar que: mide la brecha en logaritmos y presenta el valor de cada efecto en términos del logaritmo del ingreso laboral. Si se emplea la misma transformación que se usó antes en este capítulo para pasar de una escala logarítmica a una escala porcentual, se obtendrá la brecha asociada a cada componente, medida en porcentaje. Así, el efecto dotaciones equivale a un incremento del 50,2%, mientras que el efecto coeficientes (Penalización neta de la informalidad) equivale a una diferencia del 34,5%:

$$\%\Delta_dotaciones = (e^{0,4066} - 1) \times 100 = 50,2\%$$

$$\%\Delta_coeficientes = (1 - e^{-0,4233}) \times 100 = 34,5\%$$

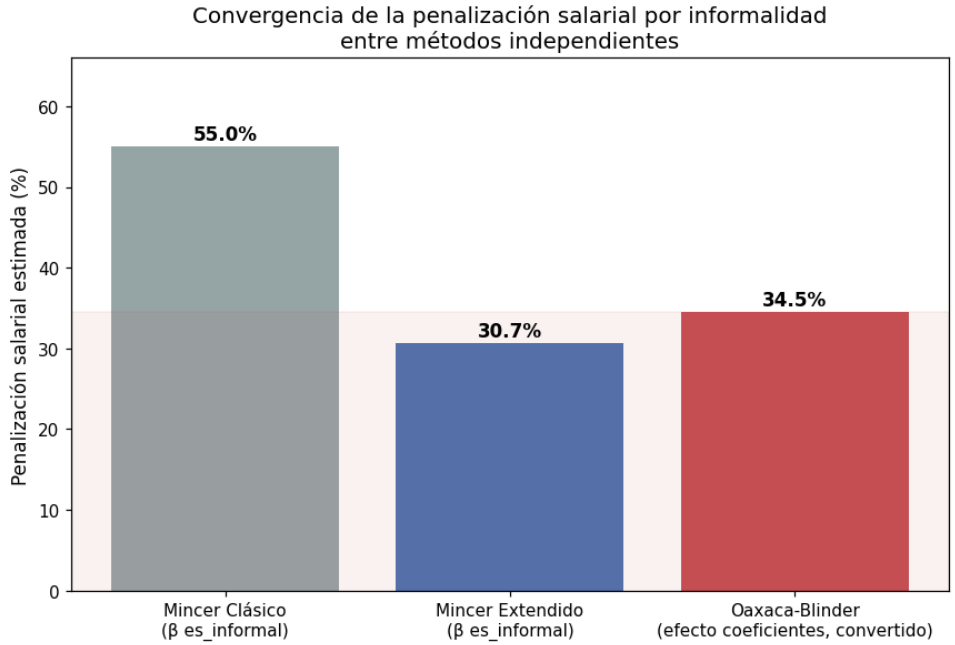


Figura 13. Convergencia de la penalización salarial por informalidad entre métodos independientes.Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.1.3.

El 34,5% no es un dato aislado. Esta cifra puede compararse, ahora sí de manera directa, con la penalización de 30,7% que arrojó la ecuación de Mincer extendida en la subsección anterior. Ambas magnitudes miden, en esencia, lo mismo: cuánto le cuesta a un trabajador ser informal, una vez descontadas las diferencias de escolaridad, experiencia, tamaño de la empresa y el resto de características observables. Sin embargo, llegan a esta conclusión por caminos completamente distintos: uno impone que formales e informales compartan la misma estructura de retornos y aísla la informalidad con una sola variable; el otro no impone nada y deja que cada grupo tenga sus propios retornos y que los datos hablen por sí solos, calculando la brecha que queda después de eso. Que dos métodos con supuestos tan diferentes coincidan en un rango tan estrecho (30,7% frente a 34,5%, una diferencia de apenas 3,8 puntos) constituye evidencia robusta de que este es el tamaño de la penalización pura asociada a la informalidad.

Finalmente, respecto a la complementariedad de los enfoques, es importante mencionar que ambos caminos converjan dentro de un rango de apenas cuatro puntos porcentuales, constituye evidencia de que la penalización salarial de la informalidad no depende de la técnica de estimación empleada. En este sentido, este ejercicio de complementariedad de modelos y equivalencia de unidades de medida ilustra, además, un aporte metodológico que merece destacarse: La descomposición de Oaxaca-Blinder combinada con Mincer Clásico y Mincer Extendido, dividiendo cada submuestra en un conjunto de entrenamiento y uno de prueba antes de calcular la descomposición final, fue una muestra del potencial de la ciencia de datos aplicada a la econometría convencional y particularmente a la economía laboral. Autores como Varian (2014) y Mullainathan y Spiess (2017) han señalado que las herramientas de aprendizaje automático y sus protocolos de validación, particularmente la separación entre entrenamiento y prueba, pueden enriquecer la práctica econométrica tradicional sin sustituirla, ofreciendo garantías adicionales sobre la capacidad de un modelo para generalizar más allá de la muestra con la que fue estimado.

Cuadro 13. Comparación de coeficientes entre la ecuación conjunta y las regresiones formal e informal separadas (en porcentaje).

Variable	**Ecuación conjunta**	**Solo formales**	**Solo informales**
Experiencia potencial (por año)	+2,2%	+1,9%	+2,2%
Escolaridad (por año)	+0,0% (n.s.)	+0,6%	-0,9%
Mujer	-30,8%	-17,3%	-42,1%
Área rural	+2,6%	-0,5% (n.s.)	+7,4%
Quintil de ingreso del hogar 2	+48,9%	+52,5%	+48,9%
Quintil de ingreso del hogar 3	+80,4%	+81,9%	+79,9%
Quintil de ingreso del hogar 4	+117,4%	+114,9%	+117,3%
Quintil de ingreso del hogar 5	+184,3%	+180,0%	+177,1%
Tamaño: microempresa 10-19 empl.	+18,9%	+21,8%	+22,9%
Ocupación: patrono	+45,2%	+41,9%	+63,2%
Ocupación: cuenta propia	-11,7%	-1,5% (n.s. al 1%)	-18,6%
Ocupación: jornalero	+13,2%	-20,6%	+11,3%
Rama: administración pública	+85,3%	+28,2%	+155,3%
Rama: comercio y vehículos	+31,9%	-2,1% (n.s.)	+45,2%

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, secciones 6.1.2 y 6.1.3.

Mientras el cuadro 12 cuantifica que el 51,0% de la brecha salarial se debe a diferencias en los coeficientes, el cuadro 13 identifica en qué variables se concentra esa diferencia, comparando una ecuación conjunta con ecuaciones separadas para el sector formal y el informal. Se destaca que el quintil de ingresos, la experiencia potencial y el tamaño de la microempresa muestran retornos similares entre ambos grupos, pero la rama de actividad, la ocupación y la escolaridad presentan retornos marcadamente distintos (por ejemplo, la escolaridad tiene un retorno positivo en el formal y negativo en el informal).

La interpretación subraya la implicación política: las intervenciones deben focalizarse en sectores y ocupaciones concretos, en lugar de aplicar incentivos uniformes. Particularmente, en el caso de la educación de la población informal, si esta no se acompaña de mecanismos de transición a la formalidad desde la política pública, la inversión en educación desde el punto de vista del retorno salarial se desperdicia, pues el sector informal no tiene cómo premiar la tecnificación, la capacitación o la educación formal de sus trabajadores. Incluso aquellos trabajadores informales que han invertido en su capital humano, al no encontrar un mejor empleo, pueden sentir que su ingreso disponible disminuyó como resultado de dicha inversión. Por ello, resultan clave las políticas de relevo entre la educación y el empleo formal.

4.3. Determinación de Perfiles de Riesgo de Informalidad Laboral

La sección anterior estableció que la informalidad impone un costo salarial cierto, de magnitud aproximada de 30 a 35 puntos porcentuales, sostenido por distintos métodos de estimación. Esta sección aborda una pregunta distinta mediante un cambio de enfoque de las brechas a la clasificación probabilística: no cuánto cuesta ser informal, sino quién tiene mayor probabilidad de llegar a serlo y cuáles son los perfiles de riesgo asociados a la informalidad. Para responder a este objetivo, resulta útil separar dos tipos de factores que la literatura y los propios modelos de esta sección distinguen con claridad: las circunstancias que una persona no eligió, es decir, su origen social, su territorio, la escolaridad de sus padres, y sobre las que la política pública tiene margen de acción directo, y las decisiones que sí se tomaron, como por ejemplo: aceptar un empleo por cuenta propia, un jornal informal, una jornada parcial, etc. Estas decisiones, a la luz de la penalización salarial ya demostrada, resultan, en la mayoría de los casos, menos rentables de lo que parecen al momento de tomarse. Para ello, la sección primero desarrolla el sustento técnico de ambos modelos (regresión logística y Gradient Boosting) y dedica la segunda mitad a traducir sus coeficientes en esa doble lectura: origen y decisión.

Determinar un perfil de riesgo exige un modelo de clasificación binaria: la variable dependiente es es_informal (1 si el trabajador es informal, 0 si es formal), la misma variable dummy construida y utilizada en la sección 4.2. Con esta información y esquema de datos, se estiman dos modelos sobre la misma población y las mismas covariables (capital humano propio y de los padres, sexo, área, estado civil, tamaño del establecimiento, grupo ocupacional, rama de actividad y quintil de ingreso del hogar) siguiendo una lógica deliberada de contraste: primero una regresión logística, que ofrece coeficientes directamente y permite evaluar la significancia estadística de cada factor; luego un Gradient Boosting, que sacrifica esa interpretabilidad directa a cambio de una mayor flexibilidad para capturar relaciones no lineales e interacciones complejas entre variables. El objetivo no es elegir un ganador, sino verificar si ambos enfoques (uno paramétrico y otro no paramétrico) convergen en señalar a los mismos determinantes como los más relevantes, lo que constituiría evidencia de que dichos factores son estructurales y no un artefacto de la técnica de estimación.

Regresión logística: planteamiento y determinantes

La regresión logística modela el logaritmo de la probabilidad de que un trabajador sea informal en relación con la probabilidad de que sea formal, como una función lineal de las covariables, de acuerdo con la siguiente forma funcional:

$$ln[p / (1 - p)] = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k$$

Donde p es la probabilidad de que el trabajador sea informal y X_1 a X_k son las covariables del modelo (escolaridad, experiencia, sexo, área, estado civil, tamaño del establecimiento, grupo ocupacional, rama

de actividad y quintil de ingreso del hogar). Al exponenciar cada coeficiente se obtiene el odds ratio, cuya interpretación es directa: valores superiores a uno indican que la característica aumenta la probabilidad relativa de ser informal, mientras que valores inferiores a uno reflejan un efecto protector frente a esa condición. Con este antecedente, el modelo se estima sobre 73.048 observaciones de entrenamiento, un tamaño muestral que otorga suficiente potencia estadística para estimar con precisión el efecto de cada factor, incluso los de menor magnitud:

Cuadro 14. Regresión logística: determinantes del riesgo de informalidad (conjunto de entrenamiento).

Variable	**Coef.**	**Odds Ratio**	**Error est.**	**z**	**P>|z|**
Constante	2,2840	9,816	0,075	30,414	<0,001
Escolaridad (year_sh)	-0,0858	0,918	0,004	-20,107	<0,001
Experiencia potencial (exp_sh)	-0,0224	0,978	0,002	-9,761	<0,001
Experiencia potencial² (exp_sh²)	0,0003	1,000	3,49e-05	9,057	<0,001
Mujer	0,1238	1,132	0,024	5,083	<0,001
Área rural	0,0339	1,034	0,029	1,172	0,241
Escolaridad del padre	-0,0076	0,992	0,003	-2,266	0,023
Escolaridad de la madre	-0,0283	0,972	0,003	-8,169	<0,001
Estado civil: casado/a	-0,5777	0,561	0,032	-17,984	<0,001
Estado civil: divorciado/a	-0,2999	0,741	0,052	-5,767	<0,001
Estado civil: separado/a	0,3310	1,392	0,045	7,421	<0,001
Estado civil: unión libre	-0,1609	0,851	0,032	-4,954	<0,001
Estado civil: viudo/a	-0,0030	0,997	0,065	-0,047	0,963
Tamaño: microempresa (10-19 empl.)	-2,3364	0,097	0,074	-31,709	<0,001
Tamaño: pequeña (20-35 empl.)	-4,0878	0,017	0,152	-26,975	<0,001
Tamaño: pequeña (36-49 empl.)	-6,4040	0,002	1,008	-6,353	<0,001
Ocupación: cuenta propia	2,5910	13,343	0,030	86,902	<0,001
Ocupación: jornalero	1,8314	6,243	0,048	37,953	<0,001
Ocupación: patrono	0,5442	1,723	0,060	9,087	<0,001
Ocupación: trabajo no remunerado	2,1970	8,998	0,152	14,486	<0,001
Rama: Administración pública	-4,9043	0,007	0,463	-10,583	<0,001
Rama: Agua y alcantarillado	-1,0281	0,358	0,299	-3,440	0,001
Rama: Alojamiento y comida	-0,9959	0,369	0,051	-19,495	<0,001
Rama: Artes y recreación	-1,2400	0,289	0,104	-11,897	<0,001
Rama: Comercio y vehículos	-1,6512	0,192	0,043	-38,298	<0,001

Rama: Construcción	0,1812	1,199	0,050	3,616	<0,001
Rama: Electricidad y gas	-1,3710	0,254	0,453	-3,025	0,002
Rama: Enseñanza	-1,4632	0,231	0,097	-15,149	<0,001
Rama: Financieras y seguros	-1,9147	0,147	0,250	-7,662	<0,001
Rama: Información y comunicación	-2,0326	0,131	0,135	-15,099	<0,001
Rama: Inmobiliarias	-1,9329	0,145	0,185	-10,424	<0,001
Rama: Manufacturas	-1,2904	0,275	0,049	-26,517	<0,001
Rama: Minas y canteras	-0,2736	0,761	0,188	-1,458	0,145
Rama: Otros servicios	-1,1782	0,308	0,061	-19,226	<0,001
Rama: Profesionales y técnicas	-2,3554	0,095	0,072	-32,778	<0,001
Rama: Salud y sociales	-1,9890	0,137	0,082	-24,383	<0,001
Rama: Servicios administrativos	-0,7356	0,479	0,074	-10,007	<0,001
Rama: Transporte	-1,2807	0,278	0,050	-25,660	<0,001
Quintil de ingreso del hogar 2	-0,4633	0,629	0,040	-11,604	<0,001
Quintil de ingreso del hogar 3	-0,8336	0,434	0,039	-21,617	<0,001
Quintil de ingreso del hogar 4	-1,1820	0,307	0,039	-30,454	<0,001
Quintil de ingreso del hogar 5	-1,7709	0,170	0,042	-41,906	<0,001

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.1.

Antes de detallar coeficiente por coeficiente —ejercicio que ya se hizo en la sección anterior para la ecuación de ingresos—, conviene identificar qué historia cuenta el conjunto de determinantes en su totalidad, porque es esa historia la que se desarrolla en el resto de esta sección. Cuatro bloques de variables emergen con el mayor peso: la posición socioeconómica del hogar de origen (el quintil de ingreso per cápita, con un gradiente protector estrictamente monótono desde el quintil más pobre hasta el más rico); el capital humano, propio y heredado (la escolaridad individual y, con menor magnitud, la de ambos padres); la escala de la unidad productiva en la que la persona se inserta (el tamaño del establecimiento, con una prima protectora que se pronuncia incluso dentro de la microempresa); y la categoría ocupacional específica que la persona ejerce (cuenta propia y jornalero, muy por delante del resto en magnitud). Los dos primeros bloques anteceden a la inserción laboral y no dependen de una decisión tomada en el mercado de trabajo; los dos últimos, en cambio, son contemporáneos a esa inserción y responden en buena medida a decisiones (o a la falta de alternativas) del propio trabajador. Esta convergencia entre determinantes estructurales y ocupacionales no es exclusiva de la ENEMDU en Ecuador: Maia de Souza et al. (2025), al clasificar trabajadores formales e informales en el área metropolitana de Monterrey mediante Gradient Boosting, encuentran que la rama de actividad, el tamaño del establecimiento y la categoría ocupacional figuran igualmente entre los predictores más relevantes, lo que sugiere que estos cuatro bloques no son una particularidad del caso ecuatoriano, sino un patrón que se repite en distintos contextos latinoamericanos.

Cuadro 15. Desempeño predictivo de la regresión logística.

Conjunto	AUC-ROC	F1	Precision	Recall
Entrenamiento	0,900	0,847	0,831	0,863
Prueba	0,897	0,838	0,817	0,861

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.1.

Respecto a las métricas del modelo, la comparación entre entrenamiento y prueba es, antes que cualquier lectura sustantiva, una prueba de sobreajuste: si el modelo memorizara particularidades de la muestra de entrenamiento en lugar de capturar una relación genuina, las métricas de prueba caerían de forma marcada respecto de las de entrenamiento. Aquí la diferencia es mínima (0,003 puntos de AUC, 0,009 de F1), lo que permite interpretar las métricas de prueba, que sí importan porque miden la capacidad del modelo sobre observaciones nunca utilizadas para ajustarlo, con confianza en que reflejan el desempeño esperado del modelo ante nuevos datos.

Sobre esas métricas de prueba: el AUC-ROC (0,897) mide la capacidad del modelo para ordenar correctamente a un trabajador informal por encima de uno formal elegido al azar. En una escala de 0,5 (equivalente a una moneda al aire) a 1,0 (clasificación perfecta); un valor de 0,897 indica una capacidad discriminante alta, muy por encima de lo que se obtendría por azar. La precisión (0,817) responde a la pregunta de cuántos de los casos que el modelo etiqueta como informales lo son efectivamente: de cada diez trabajadores que el modelo señala como informales, poco más de ocho lo son en la realidad. El recall (0,861) responde a la pregunta complementaria: de todos los trabajadores efectivamente informales, qué proporción logra detectar el modelo; en este caso, detecta a 86 de cada 100. El F1-score (0,838) resume ambas métricas en un único número mediante su media armónica, útil cuando (como en este caso) ninguna de las dos puede maximizarse sin sacrificar algo de la otra. Para un ejercicio orientado a la identificación de perfiles de riesgo con fines de política pública, un recall alto es particularmente valioso: indica que el modelo deja relativamente pocos casos de informalidad sin detectar, lo que minimiza el riesgo de que la política pase por alto a la población que busca alcanzar. Para calibrar qué tan alto es este desempeño, resulta útil compararlo con el estándar que la literatura de focalización de programas sociales utiliza para evaluar herramientas similares: Aiken et al. (2025), comparando distintos métodos para identificar hogares pobres en Bangladesh, reportan un AUC de 0,82 para la prueba de medios sustitutos basada en encuestas de hogar (investigación considerada el estándar de referencia en focalización social) frente a un AUC de 0,61. Que los modelos de esta subsección alcancen un AUC de 0,897 y 0,902 (por encima incluso de ese estándar de referencia), respalda su viabilidad como insumo para un eventual instrumento de focalización de política pública, no solo como ejercicio descriptivo.

Gradient Boosting: planteamiento y determinantes

Si la regresión logística impone una única superficie lineal sobre el log-odds, el método Gradient Boosting renuncia por completo a esa restricción; pues construye, de forma secuencial, un ensamble de árboles de decisión en el que cada árbol corrige los errores del anterior, de acuerdo con la forma general. En este sentido su forma funcional es la siguiente:

$$F(x) = \Sigma_{m=1}^{M} \gamma_m h_m(x)$$

Donde: cada $h_m(x)$ es un árbol de decisión simple ajustado sobre los errores del ensamble construido hasta el paso anterior, y γ_m es el peso con el que ese árbol se suma al resultado final; M es el número

total de árboles secuenciales, elegido en este trabajo mediante la comparación de hiperparámetros, como se describe a continuación. A diferencia de la regresión logística, esta estructura no impone una forma funcional lineal sobre el log-odds, lo que le permite capturar interacciones entre variables (por ejemplo, que el efecto protector del tamaño del establecimiento pueda ser distinto según la rama de actividad) sin necesidad de especificarlas de antemano. La selección de hiperparámetros se realizó mediante una comparación liviana entre tres profundidades máximas plausibles, evitando una búsqueda exhaustiva costosa en tiempo de cómputo: el AUC de prueba mejoró de forma consistente pero decreciente con la profundidad (0,895, 0,899 y 0,902 para profundidades 2, 3 y 4, respectivamente), por lo que se adoptó la configuración de profundidad 4.

Cuadro 16. Desempeño predictivo del Gradient Boosting.

Conjunto	AUC-ROC	F1	Precision	Recall
Entrenamiento	0,909	0,856	0,845	0,868
Prueba	0,902	0,840	0,830	0,850

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.2.

La diferencia entre entrenamiento y prueba para este modelo (0,909 frente a 0,902 de AUC) es, nuevamente, pequeña, lo que descarta un sobreajuste relevante pese a la mayor flexibilidad del modelo. En comparación con la regresión logística, el Gradient Boosting mejora el AUC de prueba en apenas 0,5 puntos porcentuales (0,902 frente a 0,897). Esa ganancia modesta es, en sí misma, un resultado sustantivo: indica que la relación entre las covariables y el riesgo de informalidad es, en su mayor parte, capturable con una especificación aditiva y lineal en el log-odds, y que el margen adicional que aporta un modelo más flexible es previsiblemente atribuible a interacciones puntuales (por ejemplo, entre grupo ocupacional y rama de actividad) más que a una reestructuración radical de la relación entre las covariables y el riesgo; ratificando la validez de las construcciones previas de esta sección y de la sección anterior. La comparación de la importancia de las variables entre ambos modelos confirma esa lectura: el grupo ocupacional, el tamaño del establecimiento, la escolaridad y el quintil de ingreso del hogar aparecen entre los determinantes más relevantes tanto en la regresión logística como en el Gradient Boosting, pese a estimarse con técnicas completamente distintas, lo que evidencia que esos determinantes son estructurales y constituyen respuestas de fondo.

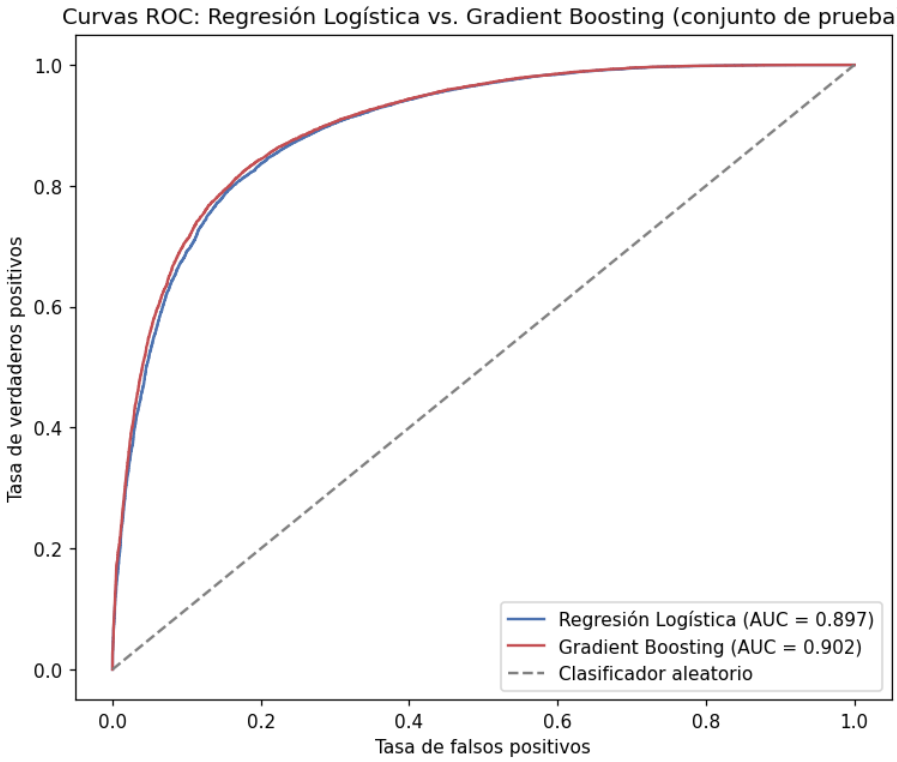


Figura 14. Curvas ROC de la regresión logística y el Gradient Boosting (conjunto de prueba).

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.3.

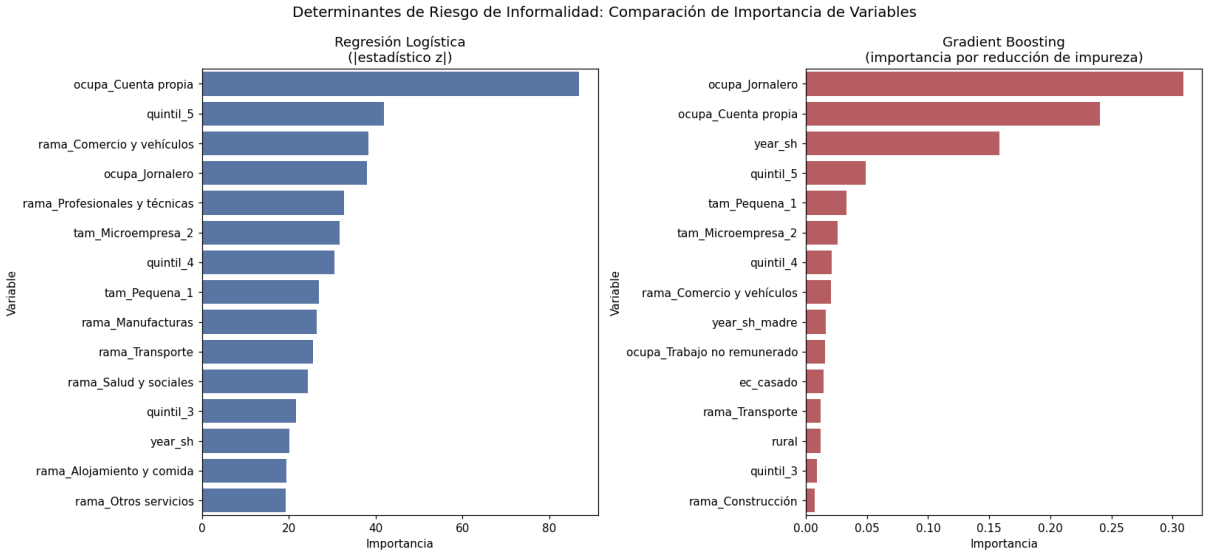


Figura 15. Comparación de importancia de variables: regresión logística vs. Gradient Boosting.

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.3.

Las Figuras 14 y 15 permiten visualizar lo que el texto ya adelantó en cifras. En la primera, las curvas ROC de ambos modelos son prácticamente indistinguibles a simple vista, reflejo directo de la diferencia de apenas 0,5 puntos en la AUC entre ellos. En la segunda, el contraste de importancia de variables confirma la convergencia de determinantes ya señalada: la cuenta propia y el jornalero encabezan ambos rankings, aunque en orden distinto (la logística prioriza la cuenta propia, el Gradient Boosting prioriza el jornalero), y el quintil de ingreso del hogar y el tamaño del establecimiento figuran entre los primeros lugares en ambos modelos. Esta coincidencia visual, lograda con dos algoritmos de naturaleza matemática completamente distinta, pero apoyados en técnicas de ciencias de datos estandarizadas, es la evidencia más directa de que los determinantes identificados no dependen de la técnica de estimación empleada.

De los coeficientes a la explicación individual: por qué recurrir a LIME

Un modelo de Gradient Boosting con centenares de árboles no ofrece, a diferencia de la regresión logística, un coeficiente único e interpretable por variable: su lógica interna combina cientos de reglas de decisión secuenciales, lo que lo convierte en una caja negra en el sentido técnico del término. Es decir, produce predicciones precisas sin que sea posible atribuir a simple vista esa precisión a razones concretas. Esta opacidad no es un problema menor cuando el objetivo último no es solo predecir, sino informar una política pública: una intervención dirigida a reducir la informalidad necesita saber no solo quién está en riesgo, sino por qué lo está, caso por caso, para poder diseñar instrumentos capaces de incidir sobre esas razones específicas. LIME (Local Interpretable Model-agnostic Explanations), resuelve esa tensión aproximando, para una observación puntual, el comportamiento del modelo complejo mediante un modelo lineal simple ajustado localmente alrededor de esa observación, recuperando así, caso por caso, una explicación tan interpretable como la de la regresión logística sin renunciar a la capacidad predictiva superior del Gradient Boosting. Las tres subsecciones siguientes se apoyan en esa técnica (y en la simulación de perfiles representativos que esta permite) para trasladar los determinantes agregados de esta subsección a historias concretas.

En síntesis, los dos modelos desarrollados en esta subsección convergen en tres conclusiones que sustentan el resto del capítulo. La primera es de robustez: una regresión logística y un Gradient Boosting (técnicas de naturaleza matemática opuesta, una lineal y paramétrica, otra no lineal y basada en árboles) coinciden en señalar a la posición socioeconómica del hogar de origen, el capital humano propio, el capital familiar heredado, la escala del establecimiento y la categoría ocupacional como los determinantes de mayor peso del riesgo de informalidad, con una ganancia de apenas 0,5 puntos de AUC a favor del modelo más flexible. Esa convergencia, visible tanto en las cifras como en la Figura 15, respalda la interpretación de que estos cuatro bloques son estructurales y no un artificio de la técnica empleada.

La segunda es de desempeño aplicado: un AUC de 0,897 a 0,902 no solo discrimina adecuadamente entre trabajadores formales e informales en términos estadísticos, sino que también supera el estándar de referencia que la literatura sobre focalización de programas sociales utiliza para evaluar herramientas comparables. Esto traslada el ejercicio de esta subsección más allá de la caracterización descriptiva: los modelos aquí estimados son, en principio, aptos para sustentar un instrumento real de identificación de la población en riesgo, y no solo para un análisis retrospectivo de sus determinantes.

La tercera, y la que organiza el resto del capítulo, es conceptual: de los cuatro bloques de determinantes identificados, dos anteceden a la inserción laboral del trabajador (su origen social y su capital humano heredado) y dos son contemporáneos a ella (la escala del establecimiento y la categoría ocupacional que efectivamente ejerce). Esta distinción entre circunstancia y decisión, más que la magnitud puntual de cada razón de momios, es la que las subsecciones 4.3.2 a 4.3.4 desarrollan en profundidad, apoyándose

en las explicaciones locales que LIME permite extraer del Gradient Boosting para traducir estos determinantes agregados en historias concretas de política pública.

4.3.2. Lo que el Origen Social Determina: los Márgenes de Acción de la Política Pública

El quintil de ingreso per cápita del hogar de origen, la escolaridad propia y materna y el área de residencia comparten una característica que ninguna categoría ocupacional comparte: anteceden por completo a la inserción laboral del individuo. Nadie elige el quintil de ingreso del hogar en el que nace, ni el nivel educativo de su madre, ni si ese hogar se ubica en una zona rural o urbana; y sin embargo, según la regresión logística de la subsección 4.3.1, estas variables se cuentan entre los determinantes de mayor peso del riesgo de informalidad que esa persona enfrentará décadas más tarde como adulto trabajador.

Esta transmisión intergeneracional de la desventaja no es un hallazgo aislado en esta investigación. Becker y Tomes (1986) formalizaron la idea de que el capital humano y las oportunidades económicas se transmiten de padres a hijos a través de la inversión familiar en educación y en el entorno de desarrollo temprano, de modo que la posición socioeconómica de origen opera como un límite —no absoluto, pero sí estadísticamente persistente— a la movilidad social de la generación siguiente. Heckman (2006) extendió este argumento al mostrar que las brechas de habilidades cognitivas y socioemocionales entre niños de distinto origen socioeconómico se abren desde la primera infancia y tienden a ampliarse, no a cerrarse, a lo largo del ciclo escolar, lo que explica por qué intervenir sobre la escolaridad del adulto -cuando esas brechas ya están consolidadas- rinde menos que intervenir sobre las condiciones del hogar de origen. Corak (2013), en su revisión de la movilidad intergeneracional de ingresos en distintos países, documenta que esta persistencia es mayor precisamente en los contextos de mayor desigualdad —la llamada curva de Gatsby—, un patrón consistente con lo ya documentado en la sección 4.1.

Para hacer tangible esta transmisión, se simulan dos pares de perfiles hipotéticos (no observaciones reales, sino combinaciones de características construidas para aislar el efecto de interés) a través del Gradient Boosting. El primer par aísla el efecto combinado del origen territorial, socioeconómico y educativo; el segundo aísla, exclusivamente, el efecto de la escolaridad materna, manteniendo todo lo demás constante.

Cuadro 17. Perfiles simulados: origen socioeconómico, territorial e intergeneracional.

Perfil	P(informal)	Ingreso esperado si formal (USD)
A1: mujer, área rural, quintil 1, educación básica	40,0%	$130
A2: mujer, área urbana, quintil 5, educación superior	4,4%	$335
B1: hombre joven, urbano, quintil 2, madre con primaria	24,7%	—
B2: igual, pero madre con educación superior	22,1%	—

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.6.

La brecha entre los perfiles A1 y A2 es de una magnitud que ningún coeficiente aislado transmite con la misma claridad: la probabilidad de informalidad del segundo perfil es una novena parte de la del primero, y su ingreso esperado más que duplica al del primero incluso antes de aplicar la penalización

salarial. Ninguna de las dos personas de este ejemplo tomó, hasta este punto de la comparación, una sola decisión ocupacional: la brecha completa se explica por dónde nacieron, en qué quintil de ingreso y con qué nivel educativo llegaron a la adultez.

El contraste entre B1 y B2 es deliberadamente más modesto (2,6 puntos porcentuales), y esa modestia es en sí misma informativa: confirma que el canal materno de transmisión intergeneracional es estadísticamente significativo y opera en la dirección esperada, pero no determina por sí solo el destino laboral de la siguiente generación; actúa en conjunto con el resto de los determinantes (quintil de origen, escolaridad propia, área), antes que sustituirlos. Tomados en conjunto, ambos pares de perfiles delinean con precisión los márgenes de acción de la política pública sobre estos determinantes: transferencias condicionadas y programas de nutrición y estimulación temprana dirigidos a hogares del quintil más bajo, con énfasis en la educación de las madres, inciden exactamente en las variables que este modelo identifica como de mayor peso entre los factores que anteceden a la inserción laboral.

4.3.3. La Decisión que Parece Rentable y no lo Es

A diferencia del origen social, la categoría ocupacional sí resulta, al menos en parte, de una decisión personal: convertirse en trabajador por cuenta propia, aceptar un jornal informal en el campo, o vincularse como asalariado en una empresa formal son, cada una, rutas que una persona puede (con distinto margen de libertad según sus circunstancias) elegir o rechazar. En el corto plazo, la primera opción puede leerse como una forma de maximizar el margen de un pequeño negocio propio sin las cargas regulatorias de la formalidad; la segunda, como una alternativa inmediata frente al desempleo abierto en zonas donde la oferta formal es escasa; y ambas, como una puerta de entrada más rápida al mercado laboral que la de esperar a una vacante formal. Lo que esta subsección muestra es que esa lectura de corto plazo, aunque comprensible, omite un costo que la sección 4.2 ya cuantificó: una penalización salarial de entre treinta y treinta y cinco puntos porcentuales, sostenida y de difícil reversión, que dicha decisión trae consigo.

Cuadro 18. Perfiles simulados: decisión ocupacional y su costo esperado.

Perfil	P(informal)	Ingreso esperado (ponderado por riesgo)
C1: cuentapropista desde los 18 años, educación media, 24 años	79,0%	$198
C2: culminó estudios superiores, empleado privado, 24 años	9,0%	$264
D1: jornalero en agricultura, rural, educación básica, 30 años	85,3%	$234
D2: asalariado privado, misma microempresa, mismo perfil	38,6%	$247

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.6.

El contraste entre los perfiles D1-D2 es el más revelador de los dos, precisamente porque es el más controlado: misma edad, mismo origen territorial y educativo, misma rama de actividad, lo único que cambia es la relación laboral específica. Y aquí aparece un matiz que conviene señalar con precisión, porque contradice la lectura más simple del problema: si ambos perfiles fueran formales, el jornalero ganaría más que el asalariado (USD 317 frente a USD 280). El trabajo de jornalero no es, en sí mismo, peor remunerado. Lo que castiga al perfil D1 no es la naturaleza de su tarea, sino la probabilidad (85,3%, más del doble que la del asalariado) de que esa tarea se realice en condiciones de informalidad, con la penalización que ello conlleva. Es, en sentido literal, el primer paso hacia el acantilado que se plantea en la apertura de esta sección: no es el salario base de la decisión, sino el riesgo que esta conlleva.

El contraste C1-C2 refuerza la misma idea desde otro ángulo, uno en el que la decisión ocupacional viene acompañada de una decisión educativa: interrumpir la formación a los dieciocho años para iniciar un emprendimiento propio, frente a completar estudios superiores antes de insertarse en el mercado laboral formal. La probabilidad de informalidad del primer camino (79,0%) es casi nueve veces la del segundo, y el ingreso esperado del segundo lo supera en un tercio pese a partir de la misma edad y del mismo punto de origen. Ni la educación ni el quintil de origen (ambos determinantes protectores y significativos según la subsección 4.3.1) logran, por sí solos, compensar el efecto de la categoría ocupacional una vez que la decisión ya fue tomada: no es posible, para quien ya es cuentapropista o jornalero informal, comprar la salida de esa condición únicamente con más años de estudio o con haber nacido en un hogar de mayores recursos. La ocupación específica que la persona termina ejerciendo domina, sobre el resto de su perfil.

4.3.4. Grupos Invisibles para la Política de Combate a la Informalidad

Las explicaciones locales de LIME sobre el Gradient Boosting, aplicadas a casos concretos del conjunto de prueba, arrojan un hallazgo no esperado: un trabajador del quintil de ingreso más alto puede presentar una probabilidad de informalidad de 0,909 (prácticamente el doble que un trabajador del quintil más pobre, con 0,462), explicada en ambos casos por la misma condición ocupacional de jornalero y diferenciada, en última instancia, por el tamaño del establecimiento en el que cada uno se inserta. Si una política de formalización dirige sus esfuerzos exclusivamente a la pobreza del hogar como proxy de vulnerabilidad laboral, este tipo de caso queda, por definición, fuera de su radar: no es pobre según el ingreso de su hogar, pero está tan expuesto a la informalidad como quien sí lo es.

Este hallazgo motivó una revisión más detallada de cómo se construye, precisamente, el quintil de ingreso que sitúa a alguien en el extremo superior de la distribución. El quintil de ingreso per cápita del hogar se calcula dividiendo el ingreso laboral del hogar entre su número de miembros, lo que implica que un mismo ingreso individual se traduce en un quintil más alto cuanto menor sea el tamaño del hogar en el que esa persona vive: quien vive solo no diluye su ingreso entre otros miembros, mientras que un jefe de hogar con un ingreso total mayor, pero repartido entre varios dependientes, puede terminar ubicado en un quintil inferior.

Cuadro 19. Composición del hogar según el quintil de ingreso per cápita.

Quintil de ingreso	Tamaño de hogar (promedio)	% hogares unipersonales
1 (más pobre)	4,61	3,3%
2	4,30	4,9%
3	3,98	7,1%
4	3,63	8,3%
5 (más rico)	2,91	17,6%

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.7.

El patrón es consistente con la hipótesis: el tamaño de hogar cae y la proporción de hogares unipersonales aumenta de forma sistemática hacia el quintil más alto. Y dentro del propio quintil 5, los hogares unipersonales muestran una tasa de informalidad de 36,2%, considerablemente mayor que la de

los hogares del mismo quintil con dos o más miembros, lo que confirma que una parte de quienes integran "el quintil más rico" no son, en rigor, personas de mayores recursos, sino individuos cuyo ingreso no se diluye entre otros miembros del hogar.

Sin embargo, es importante matizar el alcance de este hallazgo antes de generalizarlo. El ingreso laboral total del hogar —sin dividir entre sus miembros— también crece de forma monótona con el quintil (de un promedio de USD 263 en el quintil 1 a USD 1.462 en el quintil 5): los hogares del quintil más bajo no disponen, en conjunto, de más recursos que los del quintil más alto; son más grandes, pero también genuinamente más pobres en términos absolutos. Y la comparación más extrema disponible en los datos (jefes de hogar en familias numerosas (cuatro o más miembros) del quintil 1 o 2, frente a personas que viven solas en el quintil 5) muestra que los primeros presentan una tasa de informalidad de 72,8%, todavía muy superior al 36,2% de los segundos. La construcción per cápita del quintil introduce, entonces, una distorsión real pero acotada concentrada específicamente entre los hogares unipersonales del extremo superior de la distribución-, sin invalidar al quintil como medida general de la posición socioeconómica del hogar.

La implicación de política pública, en cualquier caso, se sostiene: un criterio de focalización basado únicamente en el ingreso del hogar, sea per cápita o total, deja sistemáticamente fuera a un segmento de trabajadores cuya vulnerabilidad no se origina en la pobreza de su hogar, sino en la categoría ocupacional específica que ejercen. Los hogares unipersonales de ingresos individuales moderados, insertos en ocupaciones de alto riesgo como el jornal agrícola, o cercanos a la explotación de la gig economy, constituyen precisamente ese grupo invisible: no aparecen en los registros de pobreza, pero enfrentan una probabilidad alta de informalidad y la penalización salarial que de ella se deriva, según la sección 4.2 es comparable a la de los hogares que sí aparecen.

4.3.5. Síntesis: la Matriz de Perfiles de Riesgo

Las subsecciones anteriores desarrollaron por separado el peso del origen social (4.3.2), el costo de la decisión ocupacional (4.3.3) y ejemplificaron los límites de una focalización de políticas públicas basada únicamente en el ingreso del hogar. Esta síntesis final los integra en un único ejercicio: a continuación se presenta la matriz de evaluación de perfiles de riesgo. Esta es una matriz que cruza dieciocho perfiles sociodemográficos representativos y pretende capturar la distribución nacional de tipos de trabajadores en el país. Estos perfiles se agrupan en tres macrocategorías: perfiles en vulnerabilidad, perfiles con bienestar básico y perfiles en movilidad ascendente. A su vez, estas macrocategorías fueron construidas combinando nivel de instrucción y quintil de ingreso, según se indica a continuación:

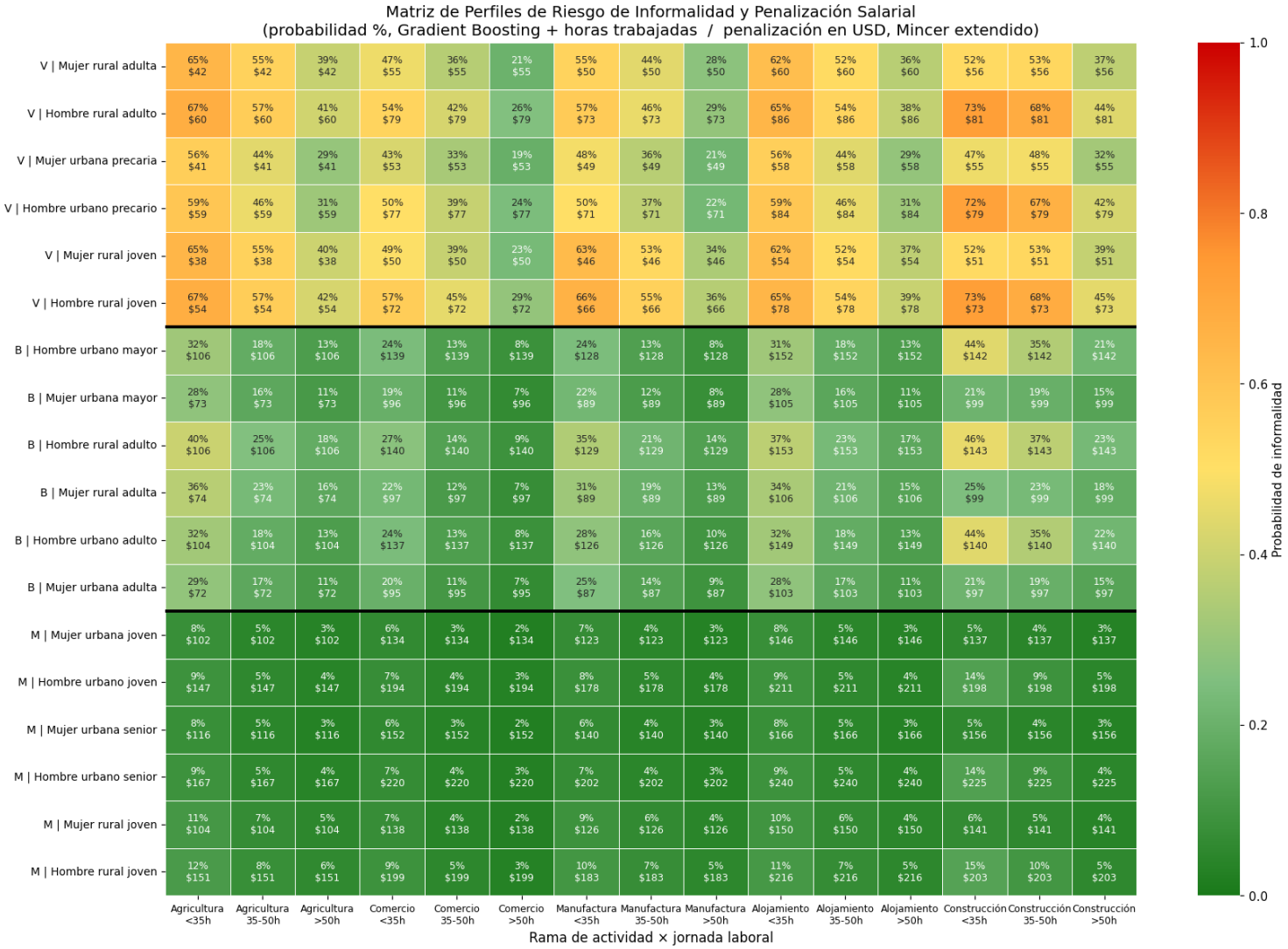
La categoría de vulnerabilidad agrupa perfiles con educación básica (7 años de escolaridad, equivalente a primaria) y que pertenecen al quintil de ingreso del hogar más bajo (quintil 1). La categoría de bienestar básico agrupa a los perfiles con educación media (doce años de escolaridad, bachillerato completo) y un quintil de ingreso intermedio (quintil 3). La categoría de movilidad ascendente agrupa a los perfiles con educación superior completa (diecisiete años de escolaridad) y el quintil de ingreso más alto (quintil 5). La elección de estas dos dimensiones —instrucción y quintil de ingreso del hogar— como base de la clasificación no es arbitraria: son, según la regresión logística de la subsección 4.3.1, los dos determinantes que anteceden a la inserción laboral del trabajador y que la sección 4.3.2 desarrolla como los principales márgenes de acción de la política pública, lo que las convierte en el criterio más natural para construir arquetipos de riesgo representativos.

Dentro de cada macrocategoría, el sexo, el área de residencia y la edad varían entre seis perfiles para capturar la heterogeneidad interna de cada grupo. Estas macrocategorías y los seis perfiles que integran cada una constituyen las filas de la matriz; mientras que las columnas de la matriz de evaluación de

perfiles de riesgo están conformadas por dos categorías: la rama de actividad y el número de horas a la semana trabajadas. Para ello, se propone como entradas (columnas) las cinco ramas de actividad mayoritariamente empleadoras; es decir: agricultura y pesca, comercio, manufactura, alojamiento y comida, y construcción.

Cada columna o rama de actividad está subdividida en tres segmentos de horas de trabajo semanales. Las categorías son: a) menos de 35 horas trabajadas a la semana; b) entre 36 y 50 horas trabajadas a la semana; y c) más de 50 horas trabajadas a la semana. Sexo, área y edad varían dentro de cada macrocategoría para reflejar la heterogeneidad interna:

Figura 16. Matriz de perfiles de riesgo de informalidad y penalización salarial.



Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: notebook adjunto, sección 6.2.8.

Según lo planteado se encuentra que:

La vulnerabilidad no se comporta como una condición aislada, sino como una acumulación: los seis perfiles de esta macrocategoría promedian 46,9% de probabilidad de informalidad, y ninguno de ellos baja de 20% incluso en su mejor combinación posible de rama y jornada —la conjunción de baja

instrucción, quintil bajo y actividad de baja productividad se refuerza en lugar de compensarse entre sí—. El bienestar básico reduce ese riesgo a 20,5% en promedio, pero de manera condicional, no garantizada: la misma persona, con la misma escolaridad y el mismo origen, enfrenta un riesgo que va de menos de 10% en comercio con jornada extensa a más de 40% en construcción con jornada corta, según dónde y cuánto trabaje. La movilidad ascendente exhibe la menor probabilidad de informalidad (5,9%) y, a la vez, la mayor penalización salarial en dólares cuando la informalidad ocurre —no una contradicción, sino el resultado esperado de aplicar un porcentaje de penalización similar sobre una base salarial más alta— y funciona como el benchmark frente al cual se miden las otras dos categorías.

Dos hallazgos transversales, presentes en las tres macrocategorías sin excepción, merecen destacarse por su valor para la política pública. El primero es que la probabilidad de informalidad decrece de forma sistemática con la jornada laboral —de 32,2% en promedio con menos de 35 horas semanales a 16,4% con más de 50—, un patrón que invierte la lectura intuitiva de que la informalidad se asocia a jornadas extenuantes: en los datos de Ecuador se asocia, con mayor fuerza, al subempleo (jornadas cortas e irregulares) que a la sobrecarga horaria, lo que sugiere que buena parte de la informalidad no es plena, sino parcial e intermitente y está estrechamente relacionada al subempleo por horas. El segundo es que la construcción resulta la rama de mayor riesgo en las tres macrocategorías sin excepción, mientras que el comercio es sistemáticamente la de menor riesgo (confirmando, con perfiles simulados y control de todas las demás características, el mismo patrón que la sección 4.1 ya había documentado con datos observados).

Conviene descender, por un momento, de los promedios por macrocategoría a las celdas individuales de la Figura 16, porque son ellas las que mejor ilustran el alcance del ejercicio. La combinación de mayor riesgo de toda la matriz corresponde a un hombre rural adulto, de baja instrucción y quintil de ingreso más bajo, empleado en la construcción con una jornada de menos de 35 horas semanales: 72,9% de probabilidad de informalidad. En el extremo opuesto, una mujer urbana mayor, con educación superior, del quintil de ingreso más alto, empleada en comercio con una jornada de más de 50 horas, presenta apenas un 2,0% de probabilidad: la distancia entre ambos extremos (casi 71 puntos porcentuales) resume, en una sola comparación, todo lo que las secciones 4.1 a 4.3.4 documentaron por separado.

El hallazgo de la jornada laboral, adelantado como patrón transversal, también se aprecia con mayor fuerza a nivel de celda que a nivel de promedio. Entre los hombres rurales jóvenes empleados en manufactura (uno de los perfiles de mayor sensibilidad a la jornada en toda la matriz), la probabilidad de informalidad cae del 66,0% con menos de 35 horas semanales al 35,8% con más de 50 horas semanales, lo que representa una reducción de 30,2 puntos porcentuales, manteniendo exactamente el mismo origen social, la misma rama y la misma edad. Que una sola dimensión (cuánto se trabaja, no dónde ni con qué origen) pueda mover el riesgo de informalidad en más de treinta puntos porcentuales dentro de un mismo perfil sociodemográfico confirma que el subempleo, y no la sobrecarga horaria, es el mecanismo que la matriz revela con mayor nitidez: la informalidad en Ecuador golpea con más fuerza a quien no logra completar una jornada plena que a quien trabaja en exceso, un matiz con implicaciones directas para el diseño de políticas de formalización que hoy se concentran, casi exclusivamente, en la regulación de jornadas excesivas.

Principales Insights de la Matriz de Evaluación de Perfiles de Riesgo para los hacedores de política pública:

1. Gradiente de riesgo según el origen social. El riesgo de informalidad cae de forma monótona y pronunciada según el origen social del perfil: de un promedio de 46,9% entre los perfiles en

vulnerabilidad a 20,5% en bienestar básico y apenas 5,9% en movilidad ascendente, lo que confirma, con perfiles simulados, el mismo patrón que la regresión logística de 4.3.1 ya había estimado con datos observados. La magnitud del salto es tan relevante como su dirección: una mujer rural adulta en situación de vulnerabilidad, empleada en la agricultura con una jornada de 35 a 50 horas, enfrenta una probabilidad del 53,4% de informalidad; una mujer urbana joven en movilidad ascendente, empleada en el comercio con esa misma jornada, enfrenta apenas 3,4%. La distancia entre ambas -cincuenta puntos porcentuales- no responde a una sola característica sino a la acumulación de todas las que distinguen a ambos perfiles.

2. Acumulación, no compensación, de desventajas —expresada como subempleo, no como sobrecarga. Ningún perfil de vulnerabilidad con riesgo bajo de 20% de riesgo, incluso en su mejor combinación posible de rama y jornada: una mujer urbana precaria, en comercio, trabajando más de 50 horas semanales— la combinación menos riesgosa disponible para todo el grupo— todavía enfrenta una probabilidad del 19,3% de informalidad. Esto confirma que la trampa de la informalidad es cierta: un perfil con déficit de educación formal y origen en el quintil más pobre no logra escapar del riesgo ni ajustando su rama de actividad ni su jornada. Conviene precisar, sin embargo, en qué consiste esa trampa: no es una trampa de sobrecarga horaria, sino de subempleo. Los datos reales de la ENEMDU muestran que estos perfiles tienen la menor probabilidad de trabajar más de 50 horas semanales de los tres grupos (3,1%, frente a 7,8% en bienestar básico y 5,7% en movilidad ascendente) y, en cambio, la mayor probabilidad de trabajar menos de 35 horas (64,8%, el doble que en los otros dos grupos). La trampa de la vulnerabilidad, entonces, no es trabajar demasiado, sino no lograr acceder a una jornada completa.

3. Condicionalidad del bienestar básico. El riesgo de este grupo intermedio no es fijo: oscila entre menos de 10% y más de 40% según la rama y la jornada —por ejemplo, un hombre urbano mayor en comercio con jornada extensa se ubica por debajo de 10%, mientras que un perfil equivalente en construcción con jornada corta supera el 40%—, lo que confirma que su posición es contingente y reversible, no una categoría estable. Con los datos reales, además, se puede afirmar con certeza algo sobre este grupo: es el que registra la mayor probabilidad de trabajar más de 50 horas semanales de los tres (7,8%, por encima del 5,7% de movilidad ascendente y muy por encima del 3,1% de vulnerabilidad). El bienestar básico no está exento de sobrecarga horaria —a diferencia de la vulnerabilidad, que está exenta de ella por subempleo—, y esa sobrecarga es probablemente parte de lo que le permite sostener su posición intermedia.

4. Jornada laboral como mecanismo transversal y quién se beneficia más de él. La probabilidad de informalidad decrece con más horas trabajadas en las tres macrocategorías, sin excepción: quienes trabajan más horas tienden a blindarse frente a la informalidad, un patrón visible tanto en construcción como en comercio y en el resto de las ramas de la matriz. Ese blindaje, sin embargo, no beneficia por igual a los tres grupos en la práctica, porque no todos tienen la misma probabilidad de acceder a una jornada extensa: movilidad ascendente y bienestar básico trabajan, en promedio, una jornada muy similar entre sí (37,8 y 37,0 horas semanales, respectivamente), mientras que vulnerabilidad promedia apenas 28,9 horas semanales. El mecanismo protector existe para los tres grupos por igual dentro de la matriz, pero solo dos de ellos —bienestar básico y movilidad ascendente— logran, en la realidad, alcanzar con frecuencia el tramo de jornada que activa ese blindaje.

5. Rama de actividad como amplificador consistente y quién se encarga de cada una. La construcción es la rama de mayor riesgo y el comercio la de menor riesgo en las tres macrocategorías sin excepción, replicando con perfiles controlados el mismo orden que la sección 4.1 documentó con datos observados. Y esto no es solo un ejercicio hipotético: con datos reales, los perfiles de

vulnerabilidad se vinculan de forma abrumadora a las ramas de mayor riesgo —59,1% se emplea en agricultura y pesca—, mientras que los perfiles de movilidad ascendente se concentran en ramas de menor riesgo que exigen mayor tecnificación, preparación o formación —42,7% en comercio, profesionales y técnicas, y salud, frente a apenas 0,4% de vulnerabilidad en ese mismo tipo de ramas—. La matriz simulada no solo describe un patrón hipotético: describe, con notable fidelidad, cómo se distribuye realmente la fuerza laboral ecuatoriana entre ramas de distintos niveles de riesgo.

6. Disociación entre riesgo y penalización -posiblemente la conclusión más relevante de la matriz. Un hombre urbano senior en movilidad ascendente, empleado en alojamiento y comida, enfrenta una probabilidad de informalidad apenas del 5,2%, pero una penalización salarial de USD 239,66 si esa informalidad llegara a ocurrir. Una mujer rural joven en situación de vulnerabilidad, empleada en la agricultura, enfrenta una probabilidad del 54,7% de informalidad —diez veces más—, pero una penalización de apenas USD 37,63 si ocurre —seis veces menos—. El riesgo y la penalización no se mueven en la misma dirección: quien tiene menor probabilidad de caer en la informalidad es, paradójicamente, quien más perdería si cayera, porque parte de una base salarial mucho más alta. Ambos ejes deben leerse juntos: un perfil de bajo riesgo y alta penalización potencial no es lo mismo que uno de alto riesgo y baja penalización potencial, y una política de formalización que solo mire el riesgo, sin considerar la penalización, ignoraría la mitad de la historia económica de cada perfil.

7. Especialización sectorial real por macrocategoría. Con datos observados —no simulados—, la vulnerabilidad se concentra en la agricultura (59,1%) y la movilidad ascendente en ramas tecnificadas (42,7%), lo que confirma que la matriz simulada no describe una segmentación arbitraria, sino que refleja, con notable precisión, una segmentación productiva genuina del mercado laboral ecuatoriano.

8. Subempleo como expresión real de la precariedad. 64,8% de los trabajadores en situación de vulnerabilidad trabajan menos de 35 horas semanales, el doble que en los otros dos grupos, y ese déficit de horas se traduce directamente en un déficit de ingresos: el perfil promedio de vulnerabilidad en la matriz predice ingresos apenas entre USD 38 y USD 84, según la combinación de rama, frente a cifras entre USD 100 y USD 280 en movilidad ascendente. La combinación de pocas horas y un bajo ingreso por hora —no solo la informalidad en sí misma— es la que define la precariedad de este grupo: con la mayor probabilidad de los tres grupos, son quienes caen en un déficit simultáneo de horas trabajadas e ingresos suficientes.

9. Validación cruzada entre secciones del capítulo. Los patrones de la matriz -gradiente de riesgo por origen social, efecto protector de la jornada, amplificación por rama de actividad- coinciden, punto por punto, con los hallazgos de las secciones 4.1, 4.2 y 4.3.1 a 4.3.4, pese a haberse obtenido con técnicas y fuentes de datos distintas -estadística descriptiva, regresión logística, Gradient Boosting, y ahora una matriz de perfiles simulados-. Esta convergencia reiterada es, en sí misma, evidencia de que las conclusiones de este capítulo son estructurales y no casuísticas: no dependen de una técnica particular ni de un corte específico de los datos, sino que emergen una y otra vez bajo enfoques metodológicos independientes.

10. Aplicabilidad directa, incluida su dimensión vocacional. La matriz permite anticipar, para una combinación específica de origen social, rama de actividad y jornada laboral, tanto la probabilidad de que esa persona termine en la informalidad como el costo salarial que esa condición le impondría —un insumo operativo para la focalización de políticas públicas, no solo un diagnóstico descriptivo—. Pero su utilidad no se agota en el diseño de política: también puede funcionar como una herramienta de orientación vocacional para quienes están por tomar decisiones laborales o iniciar su vida laboral, permitiéndoles identificar en qué tipo de perfil se ubican hoy o en cuál aspiran a ubicarse, y anticipar,

con datos —no solo con intuición—, las recompensas y penalizaciones que el mercado laboral ecuatoriano les depara según esa elección.

Esta sección desarrolló tres piezas de evidencia que, tomadas en conjunto, delinean un mismo diagnóstico. La informalidad laboral en Ecuador no es aleatoria en su origen: se concentra de manera sistemática entre quienes nacieron en hogares de menores recursos, con madres de menor escolaridad y en zonas rurales, circunstancias que ninguna de esas personas eligió (secciones 4.1 y 4.3.2). Tampoco es gratuita en su ejercicio: quienes la asumen como decisión de corto plazo (por cuenta propia, como jornaleros, en jornadas parciales) enfrentan una penalización salarial cierta, de treinta a treinta y cinco puntos porcentuales, que ni la educación ni el origen socioeconómico logran amortiguar una vez tomada la decisión (secciones 4.2 y 4.3.3). Y su combate exige mirar más allá de quién es pobre, hacia quién está ocupacionalmente atrapado: un criterio de focalización basado solo en el ingreso del hogar deja sistemáticamente fuera a trabajadores (frecuentemente en hogares pequeños o unipersonales) cuya vulnerabilidad se origina en la ocupación que ejercen y no en la pobreza que declaran. La matriz de perfiles de riesgo de esta última subsección ofrece, en ese sentido, una herramienta concreta: permite anticipar, para una combinación específica de origen social, sector económico y jornada laboral, tanto la probabilidad de que esa persona termine en la informalidad como el costo salarial que esa condición le impondría, insumo directo para dirigir con mayor precisión los instrumentos de formalización hacia los perfiles y las ramas donde el riesgo y la penalización que lo acompaña, son más altos.

Conclusiones

A partir de los estándares de la Conferencia Internacional de Estadísticos del Trabajo de la OIT (17.ª y 19.ª CIET), la investigación adoptó un enfoque legalista, distinguiendo con claridad la dimensión del Sector Informal (características de la unidad económica) frente a la del Empleo Informal (condiciones del puesto de trabajo). La fundamentación en el Sector Informal ofrece criterios objetivos, verificables y fiscalmente operativos (registro tributario, contabilidad organizada y tamaño del establecimiento), superando las limitaciones del enfoque productivista tradicional al medir la formalización de la estructura productiva y relacionándola con el tamaño del establecimiento, la ausencia de capital y el autoempleo.

Asumiendo este enfoque, el diagnóstico empírico se basó en los microdatos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU - INEC, 2025), aplicando los factores de expansión poblacional correspondientes para reflejar la estructura de una Población Económicamente Activa (PEA) de 8.612.073 personas. Según la definición legalista, la tasa de informalidad en Ecuador se ubicó en el 49,65% de la PEA (4.275.784 personas). No obstante, el enfoque legalista tiene matices e incluye una versión suave de medición y otra más reforzada. Bajo la versión reforzada de este enfoque, además de considerar las categorías tomadas en cuenta para la medición de la informalidad, habría que incorporar dentro del análisis el acceso a la seguridad social pública o privada y la existencia de estabilidad laboral mediante la existencia de contratos firmados. En este sentido, con la incorporación únicamente de la ausencia de afiliación a la seguridad social, la tasa podría superar fácilmente el 60% de la PEA.

Por otro lado, la definición adoptada tampoco incluye en su universo de estudio al empleo doméstico no afiliado (163.817 trabajadores pertenecientes a la economía transaccional de consumo), lo que elevaría la tasa real del sector informal al 52%. Como advierten varios autores en el marco metodológico aquí presentado, el empleo doméstico pertenece a la economía transaccional y no al autoconsumo de los hogares; por lo tanto, no existe razón técnica para aislarlo de la contabilidad del sector informal. Sin embargo, se excluye de la presente aplicación de este trabajo para lograr la comparabilidad con la metodología oficial.

En consecuencia, se advierte una posible subestimación del tamaño del sector informal. Esta, más allá de ser una precisión técnica sobre la aplicación del presente estudio, constituye una alerta respecto a la construcción de la metodología oficial de empleo adoptada por el Instituto Nacional de Estadística y Censos. Sin embargo, aun con este subdimensionamiento presente, la tasa de informalidad refleja la dimensión de la informalidad laboral en el país e ilustra claramente la brecha que presenta el mercado laboral ecuatoriano en términos de la disparidad de bienestar de sus trabajadores. Uno de cada dos ocupados en el país está vinculado a prácticas laborales precarias, que lo alejan de su ingreso laboral potencial y, cada día en esa condición, inciden en su bienestar presente y futuro.

Una vez dimensionado este fenómeno, el estudio presentó una tipología funcional para comprender su composición (informalidad asalariada, cuentapropista y no remunerada). La evidencia desarrollada presentada demostró que la informalidad se distribuye de forma marcadamente desigual entre tipologías ocupacionales, territorios y niveles educativos: tal es el caso que la informalidad por cuenta propia concentra el núcleo más numeroso del fenómeno (27,37% de la PEA, equivalente a 2.356.496 personas), seguida por la informalidad asalariada (13,51%) y por el trabajo no remunerado (8,78%), siendo esta una composición que desmiente cualquier lectura uniforme del problema. En este sentido, el presente estudio invitó al lector a dejar de hablar del sector informal y a hablar del espectro informal y sus perfiles.

No existe una única informalidad, sino al menos tres experiencias distintas de inserción precaria en el mercado de trabajo, cada una con determinantes y consecuencias propias.

Con esta perspectiva, se analizó la dinámica interna de quienes son informales a la luz de sus características socioeconómicas y decisiones de vida: La sección de caracterización de la informalidad permitió arrojar las siguientes conclusiones:

1. **La educación como escudo: el gradiente educativo inverso:** La caracterización de este estudio evidencia que la escolaridad funciona como un amortiguador de capital humano frente a la informalidad: a mayor nivel educativo, la probabilidad de permanecer en empleos informales disminuye de manera sostenida, mientras que los niveles educativos más bajos concentran tasas extremas de informalidad. Puntualmente, la tasa de informalidad alcanza el 87,45% entre quienes no cuentan con ningún nivel de instrucción y el 93,85% entre quienes registran educación básica como máximo nivel alcanzado". Esto contrasta con las tasas de informalidad de la población con educación superior universitaria, superior no universitaria o de posgrado, pues en ningún caso supera el 21%. Estos resultados obligan a distinguir entre dos objetivos de política: por un lado, ampliar la cobertura educativa para reducir la exposición inicial a la informalidad; por otro, mejorar la calidad y la pertinencia de la formación para que los años de estudio se traduzcan efectivamente en empleos formales.

2. **La trayectoria en forma de U de la informalidad en la dinámica intergeneracional: Se evidenció que las tasas de informalidad a lo largo de la distribución etaria de la población presentan** una trayectoria en U. Es decir, las tasas de informalidad son particularmente elevadas entre jóvenes y adultos mayores, y menores en los tramos de edad económicamente más productivos, lo que sugiere que la informalidad funciona como una alternativa precarizadora frente al desempleo juvenil y como refugio de subsistencia para permanecer dentro de la PEA en la vejez. Esta dinámica tiene implicaciones distintas según la etapa de vida: para los jóvenes, requiere políticas de inserción laboral y formación técnica vinculada a la demanda; para los adultos mayores, redes de protección social y mecanismos de transición.

3. **La brecha de género asociada a la informalidad:** Paralelamente, aunque el estudio no busca implementar un análisis de género per se, evidenció que el mercado laboral al día de hoy continúa imponiendo un castigo a las mujeres: Al observar cómo los ingresos evolucionan a lo largo del ciclo de vida, diferenciando por género y condición de formalidad; se encontró que tanto los hombres que mantienen un empleo formal, como los hombres que mantienen un empleo informal, en promedio perciben ingresos superiores a mujeres con empleos formales a lo largo de todo el ciclo de vida. Adicionalmente, la trayectoria de ingresos de los hombres es cóncava, lo que evidencia una movilidad ascendente y mayores ingresos a medida que se acercan a sus años de mayor productividad. Sin embargo, la trayectoria de ingresos de las mujeres, sobre todo de las informales, no es cóncava sino lineal. Es decir, conforme se acercan a sus años de mayor productividad, sus ingresos laborales no necesariamente suben, por lo que se presupuesta que el ingreso con el que cuentan al insertarse por primera vez en el mercado laboral será el mismo ingreso o muy cercano al que alcancen en sus años de mayor productividad. Ello merece un zoom óptico mayor; sin embargo, la estabilidad, continuidad e incremento de sus ingresos se ven restados debido a las labores de cuidado que ejercen y a que en reiteradas ocasiones durante su vida laboral deben interrumpir sus trabajos para dedicarse a la economía del cuidado en el hogar. Estas desconexiones o inactivaciones recurrentes del mercado laboral (por ejemplo, por permiso de maternidad), dentro de un esquema de

informalidad, se penalizan con desempleo o con la subvención del trabajador sustituto, lo que afecta negativamente al ingreso disponible.

4. **La heterogeneidad sectorial de la informalidad:** Se demostró que la informalidad tampoco se distribuye de manera uniforme entre las ramas de actividad económica, sino que se concentra en las menos intensivas en mano de obra calificada. Puntualmente, se observó que Agricultura, Ganadería y Pesca concentran el mayor volumen de informalidad del país, dominado por el cuentapropismo de subsistencia y el trabajo no remunerado. Complementariamente, sectores como el comercio presentan una alta densidad de informales asalariados, reflejo de la baja barrera de entrada y de la proliferación de microunidades de venta y de servicios que operan sin regulación. De la misma forma, Alojamiento y Servicios de Comida también figuran entre las actividades mayoritariamente informales. El denominador común de estas ramas de actividad es la baja tecnificación de su fuerza laboral (lo que funciona como puerta de entrada para que trabajadores con déficit de formación, capacitación o educación se integren). CEPAL (2020) las identifica como actividades de "baja productividad", en las que la informalidad no responde a una elección individual sino a una estrategia de supervivencia. Mención específica corresponde a la gig economy, y a la economía de las plataformas; que pese a que no constituyen una rama de actividad específica, atraen debido a la desregulación existente, a trabajadores informales que encuentran en ellas una nueva alternativa de supervivencia; sin embargo, este fenómeno con la desagregación actual de la encuesta de empleo no pude ser estudiado de cerca.

5. **La brecha de ingresos:** una primera aproximación monetaria con implicaciones de política. El análisis descriptivo demostró que existe una penalización salarial sustantiva asociada a la informalidad: la brecha monetaria asociada a esta condición, documentada de forma preliminar y descriptiva en esta primera etapa, ya insinuaba su magnitud: un trabajador formal percibe, en promedio, 3,13 veces el ingreso de uno informal (621,85 dólares frente a 198,44 dólares). Sin embargo, esta es una diferencia que la estadística descriptiva y el análisis exploratorio de datos todavía no permitían atribuir con precisión a la condición de informalidad en sí misma.

Con el fin de determinarlo, se pusieron a prueba la econometría convencional y la ciencia de datos para verificar si sus enfoques arrojaban resultados equivalentes que permitieran contarle a un trabajador informal cuánto de su ingreso está dejando de recibir por permanecer aislado de la formalidad, aun cuando en el corto plazo le parezca que está ahorrando costos operativos y le evite incurrir en costos de formalización.

Para ello, se establecieron tres momentos: a) la aplicación de la ecuación de Mincer en su forma clásica, b) la aplicación de la ecuación de Mincer en su forma extendida; y c) un enfoque más cercano a la ciencia de datos mediante la formulación del modelo de Oaxaca-Blinder. La estimación de la penalización salarial confirmó que esa brecha bruta inicialmente identificada, aunque real, sobreestima el efecto neto de la informalidad al no distinguir entre lo que es penalización propia asociada a esa condición y lo que responde al déficit, al superávit o a la variación de dotaciones o características asociadas a los trabajadores informales.

En este sentido, la ecuación de Mincer clásica, que no incorpora controles estructurales, arrojó una penalización global del 55,0% entre trabajadores informales y trabajadores formales equivalentes en dotaciones o características. Al refinar el planteamiento de Mincer Clásico y adaptarlo al caso ecuatoriano, incorporando variables de interés tales como el tamaño del establecimiento, la rama de actividad, la categoría ocupacional y el quintil de ingreso del hogar, la ecuación extendida redujo esta

penalización a 30,7%, todavía negativa y significativa, pero con casi la mitad de su magnitud original absorbida por esas variables estructurales antes omitidas. La descomposición de Oaxaca-Blinder, aplicada como validación metodológicamente independiente de ese resultado, corroboró el hallazgo desde un ángulo distinto: 51,0% de la brecha salarial entre formales e informales no se explica por diferencias en dotaciones observables, sino por el hecho de que el mercado retribuye esas mismas características de manera distinta según la condición de formalidad del trabajador, una cifra que, convertida a una escala porcentual comparable con la de la ecuación de Mincer extendida, equivale a 34,5%. Siendo esta una diferencia de apenas 3,8 puntos porcentuales entre dos métodos que parten de supuestos opuestos sobre la estructura de retornos del mercado laboral, y nacen desde perspectivas diferentes. La una, siendo insignia de la econometría clásica y de la economía laboral; y la otra, proveniente de la vertiente de la ciencia de datos.

Esta convergencia constituye evidencia sólida de la existencia de una penalización cierta asociada a la informalidad. Debido a la cercanía de las estimaciones al emplear métodos distintos, se confirma que la misma no es un artificio producto de la técnica de estimación empleada, sino una regularidad del mercado laboral ecuatoriano que persiste bajo enfoques metodológicos independientes. Complementariamente, este hallazgo permite confirmar que la aplicación del método KDD (Knowledge Discovery in Databases), (pese a que no se lo ha hecho explícito en este documento, sino en el Anexo 1 (cuaderno de trabajo adjunto), específicamente en la introducción de la sección 4 (Preprocesamiento y Transformación de Datos), donde se detallan las cuatro fases del método y su correspondencia con cada sección del cuaderno); contribuye a minimizar el ruido en los datos y a mejorar la precisión de las estimaciones, dado que el proceso sistemático de selección, limpieza, transformación y validación favorece la obtención de insights consistentes entre métodos.

El 49,0% restante de la brecha total sí respondió a dotaciones, es decir, a que, en promedio, las formales e informales difieren en sus características observables. El componente de coeficientes (el 51,0% recién descrito) reveló, además, que la penalización no es uniforme entre variables: mientras que el quintil de ingreso del hogar de origen se retribuye de forma prácticamente idéntica entre formales e informales, la rama de actividad, la ocupación y la escolaridad exhiben retornos marcadamente distintos entre ambos grupos. La escolaridad ofrece, en este sentido, el ejemplo más elocuente de por qué el problema no se agota en las dotaciones: incluso entre trabajadores con exactamente el mismo nivel educativo, el mercado los retribuye de forma opuesta según su condición de formalidad —dentro del sector formal, cada año adicional de estudio se asocia con un aumento de 0,6% en el ingreso; dentro del sector informal, con una disminución de 0,9%—. Esta no es solo una diferencia de magnitud: el signo se invierte: estudiar más ayuda si el trabajador es formal, pero no contribuye (incluso perjudica ligeramente, una vez controlado todo lo demás) si el trabajador es informal. De ello específicamente se deriva una conclusión importante: la política de capacitación y educación no debe diseñarse ni evaluarse de forma aislada, tiene que ir acompañada de una puerta de transición hacia la formalidad, o se desperdicia. Si el Estado invierte en capacitar a un trabajador informal, pero esa persona permanece en el sector informal después de la capacitación, la evidencia indica que esa inversión no se traduce en un mayor ingreso; incluso puede coincidir con una ligera penalización. El problema no es la capacitación en sí, sino que el trabajo informal no tiene la estructura (contratos, categorías, jerarquías salariales ligadas a credenciales) para pagarla.

De ahí se derivan dos implicaciones concretas de política. En primer lugar, los programas de capacitación dirigidos a la población informal deberían tener, como componente obligatorio, un mecanismo de tránsito hacia la formalidad, tales como bolsas de empleo formal, convenios con

empleadores formales, certificación reconocida por el sector formal, en lugar de ser programas de formación aislados sin conexión con la demanda formal de trabajo.

En segundo lugar, reforzar la política de formalización no es opcional frente a la política educativa, es su complemento necesario: capacitar sin abrir la puerta de entrada a la formalidad es empujar a más gente a competir por los mismos puestos informales que ya no pueden pagar mejor por más estudios o mayor preparación. Y es que resulta obvio, si un trabajador gasta de su ingreso disponible para prepararse, por ejemplo haciendo una carrera técnica, un diplomado, obteniendo una certificación o una maestría, pero el mercado formal no le abre las puertas; al final del día su retorno de la educación es negativo. En este sentido, pese a que formalmente el desempleo no ha sido incorporado como variable de análisis, el modelo capta bastante bien la ausencia generalizada de plazas formales de trabajo, lo que se traduce en que los trabajadores informales capacitados perciban retornos negativos asociados a su capacitación, tecnificación o educación formal.

En síntesis: este hallazgo es evidencia adicional de que la política pública debe atacar la oferta (educación) y la demanda (formalización) a la vez, porque, sin la otra, el capital humano invertido quedará sin ningún retorno que cobrar.

El tercer bloque de resultados desplazó la pregunta de investigación desde la magnitud de la penalización hacia la identificación de qué perfiles enfrentan mayor probabilidad de caer en la informalidad, mediante un ejercicio de clasificación que combinó regresión logística y Gradient Boosting. Ambas técnicas, una paramétrica y lineal, otra no paramétrica y basada en árboles de decisión, convergieron en un desempeño predictivo alto (AUC de 0,897 y 0,902, respectivamente) y en señalar a los mismos cuatro bloques de determinantes: dos que anteceden a la inserción laboral del trabajador (su origen socioeconómico y su capital humano, propio y el capital humano heredado de sus padres) y dos que son resultado de ella (la escala del establecimiento y la categoría ocupacional que efectivamente ejerce). Esta distinción entre circunstancias heredadas y decisiones implementadas organizó, a su vez, la construcción de dieciocho perfiles representativos agrupados en tres categorías: vulnerabilidad, bienestar básico y movilidad ascendente.

Los perfiles de vulnerabilidad no solo concentran la mayor probabilidad de informalidad (46,9% en promedio), sino que revelan un patrón novedoso de subempleo estructural: su riesgo decrece con el incremento de las horas laborables a la semana, pero rara vez alcanzan jornadas completas. Cabe destacar que 6 de cada 10 de estos perfiles se insertarán en la agricultura (59,1%), lo que evidencia que su exclusión no es solo salarial, sino también horaria y sectorial. Los perfiles en bienestar básico, en cambio, no ocupan una posición de riesgo intermedio fijo, sino una posición profundamente condicional: su probabilidad de informalidad oscila con fuerza según la rama y la jornada específicas en las que se insertan, lo que sugiere que este grupo constituye la población con mayor margen de maniobra para la política pública, al no estar aún atrapado en un patrón estructural único. Los perfiles en movilidad ascendente, finalmente, no solo exhiben el menor riesgo (5,9%) de caer en la informalidad, sino que su inserción en ramas que exigen tecnificación (42,7%) confirma que su protección frente a la informalidad también es sectorial.

En conjunto, la matriz de perfiles de riesgo constituye, más allá de sus hallazgos específicos, un instrumento cuyo potencial trasciende el diagnóstico de política pública: aunque este trabajo la desarrolla como un esfuerzo inicial, su lógica de anticipar, para una combinación dada de características, la probabilidad de informalidad y el costo salarial asociado, para cada rama de actividad en la que el trabajador quiera insertarse, sienta las bases para su eventual uso en ejercicios de orientación vocacional,

permitiendo a una persona visualizar con evidencia empírica las consecuencias esperables de distintas trayectorias ocupacionales antes de emprenderlas.

Leídos en conjunto, los tres bloques de resultados del presente trabajo convergen en un diagnóstico único, sustentado por métodos independientes y fuentes de evidencia complementarias. La informalidad laboral ecuatoriana no es aleatoria en su origen, se concentra de manera sistemática entre quienes nacieron en hogares de menores recursos, con madres de menor escolaridad y en zonas rurales, circunstancias que ninguna de esas personas eligió; ni gratuita en su ejercicio: quienes la asumen como decisión de corto plazo enfrentan una penalización salarial cierta, de entre treinta y treinta y cinco puntos porcentuales Su combate e instrumentación, en consecuencia, exige mirar más allá de quién es pobre, hacia quién está ocupacionalmente atrapado: los instrumentos de política pública orientados exclusivamente al capital humano, tales como programas de capacitación, o nivelación educativa, inciden sobre poco menos de la mitad del problema, mientras que la porción mayor, aquella que responde a cómo el mercado retribuye de forma distinta a formales e informales con las mismas características, demanda instrumentos de otra naturaleza: regímenes simplificados de registro diseñados específicamente para las actividades donde la divergencia de retornos es más severa (el comercio informal urbano y el jornal agrícola y de construcción, por ejemplo); en lugar de un régimen único aplicado a toda la economía informal; esquemas de afiliación a la seguridad social con aporte diferenciado o subsidiado para trabajadores por cuenta propia, que hoy pagan 17,60% de su ingreso declarado en su totalidad, frente al 9,45% que asume un asalariado formal mientras su empleador cubre el resto, y mecanismos de registro simplificado para empleadores ocasionales de jornaleros, que hoy no tienen ningún incentivo administrativo para reportar contrataciones de pocos días; cruce de información entre el Servicio de Rentas Internas, el IESS y los gobiernos autónomos descentralizados, de modo que la informalidad deje de ser una opción sin costo administrativo detectable; y tarifas diferenciadas en servicios básicos (agua, electricidad, telecomunicaciones) condicionadas a la formalidad del establecimiento, que trasladen el beneficio fiscal de un subsidio generalizado hacia un incentivo explícito a la formalización.

Recomendaciones

De los hallazgos anteriores se desprenden recomendaciones que, en conjunto, exigen actuar sobre tres frentes simultáneos: la medición oficial de la informalidad, sus determinantes estructurales de origen y la estructura de incentivos que hoy desalienta la transición hacia la formalidad.

En materia de medición, se recomienda al Instituto Nacional de Estadística y Censos evaluar incorporar el empleo doméstico no afiliado al universo de cómputo del sector informal, dado que, según se documentó, no existe sustento técnico para excluirlo (pertenece a la economía transaccional, no al autoconsumo del hogar) y su exclusión eleva artificialmente la cifra oficial en al menos tres puntos porcentuales de la PEA. Del mismo modo, se recomienda complementar la medición legalista vigente con reportes periódicos en su versión reforzada (incorporando la afiliación a la seguridad social), aun cuando no se adopte como definición oficial, de manera que la política pública cuente con ambos extremos del fenómeno y no solo con su versión más conservadora.

En materia de determinantes de origen, la evidencia de un gradiente educativo inverso y de una trayectoria etaria en forma de U exige diferenciar la política según la etapa de vida: programas de inserción laboral y formación técnica vinculada a la demanda para la población joven, y mecanismos de protección social y de transición gradual para la población adulta mayor, en lugar de un único instrumento aplicado indistintamente a ambos grupos. La persistencia de una brecha de género que se amplía (no se cierra) a lo largo del ciclo de vida laboral recomienda, además, incorporar explícitamente servicios de cuidado y mecanismos de reinserción posteriores a interrupciones laborales en el diseño de cualquier programa de formalización, para evitar que estos programas reproduzcan la misma penalización que buscan corregir.

En materia de incentivos, y dado que el 51,0% de la brecha salarial responde a coeficientes (a cómo el mercado retribuye de forma distinta a formales e informales con las mismas características) y no a un déficit de capital humano, se recomienda priorizar instrumentos que reduzcan directamente el costo de formalizarse, por encima de instrumentos que solo eleven las calificaciones de la fuerza laboral informal. Entre estos, resultan prioritarios: regímenes simplificados de registro dirigidos específicamente a las ramas y ocupaciones donde la divergencia de retornos es más severa (comercio informal urbano, jornal agrícola y de construcción); esquemas de afiliación a la seguridad social con aporte diferenciado o subsidiado para trabajadores por cuenta propia, que hoy asumen 17,60% de su ingreso declarado en su totalidad frente al 9,45% que asume un asalariado formal; mecanismos de registro simplificado para empleadores ocasionales de jornaleros; cruce sistemático de información entre el Servicio de Rentas Internas, el IESS y los gobiernos autónomos descentralizados; y tarifas diferenciadas en servicios básicos condicionadas a la formalidad del establecimiento. Todo programa de capacitación dirigido a población informal debería, además, incorporar como componente obligatorio un mecanismo de tránsito hacia la formalidad (bolsas de empleo formal, convenios con empleadores formales, certificación reconocida por el sector formal), dado que la evidencia muestra que la capacitación sin puerta de salida hacia la formalidad no se traduce en mayor ingreso, e incluso puede coincidir con una penalización adicional.

Finalmente, como línea de investigación futura, se recomienda explorar la desagregación específica de la gig economy y la economía de plataformas dentro de la encuesta de empleo, actualmente no identificable con el nivel de detalle de la ENEMDU, así como extender la matriz de perfiles de riesgo hacia una herramienta de orientación vocacional que permita a un trabajador, estudiante o persona económicamente inactiva a punto de iniciar su incursión en el mundo laboral, antes de tomar una decisión ocupacional, anticipar con evidencia empírica sus consecuencias esperables en términos de riesgo de informalidad y penalización salarial.

Referencias Bibliográficas

Aiken, E., Ashraf, A., Blumenstock, J., Guiteras, R., y Mobarak, A. M. (2025). *Scalable targeting of social protection: When do algorithms out-perform surveys and community knowledge?* (NBER Working Paper No. 33919). National Bureau of Economic Research.

Argueta, S. y Li, C. (2026). *América Latina en la era inteligente.* Foro Económico Mundial.

Bairoch, P. (1971). *Les écarts des niveaux de développement économique entre pays développés et pays sous-développés de 1770 à 2000* [Las diferencias en los niveles de desarrollo económico entre países desarrollados y subdesarrollados de 1770 a 2000]. *Revue Tiers Monde, 12*(47), 497–514.

Banco Interamericano de Desarrollo. (2015). *Informe sobre desarrollo en América Latina: Paredes de cristal y brechas de género.* BID.

Bangasser, P. E. (2009). *The ILO and the informal sector: An institutional history.* International Labour Office.

Becker, G. S. y Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics, 4*(3), S1–S39.

Benton, L. (1989). Industrial subcontracting and the informal sector: The politics of restructuring in the Madrid electronics industry. En A. Portes, M. Castells y L. Benton (Eds.), *The informal economy: Studies in advanced and less developed countries.* The Johns Hopkins University Press.

Blau, F. D. y Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature, 55*(3), 789–865.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources, 8*(4), 436–455.

Card, D. (1999). The causal effect of education on earnings. En O. Ashenfelter y D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1801–1863). North-Holland.

Comisión Económica para América Latina y el Caribe. (2020). *Panorama social de América Latina 2020.* CEPAL.

Comisión Económica para América Latina y el Caribe. (2024). *Panorama fiscal de América Latina y el Caribe 2024.* CEPAL.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives, 27*(3), 79–102.

De Soto, H. (1989). *El otro sendero.* Editorial Oveja Negra.

Dierckxsens, W. (1989). *Mercado de trabajo y política económica en América Central.* Universidad Nacional de Costa Rica.

Doeringer, P. y Piore, M. (1971). *Internal labor markets and manpower analysis.* Heath.

Enríquez Sánchez, O. L. (2014). *El reto de la política fiscal en la informalidad.* Repositorio Digital PUCE.

Foro Económico Mundial. (2025). *Informe sobre exclusión tecnológica y el futuro del trabajo en el Sur Global.* FEM.

Gasparini, L. y Tornarolli, L. (2010). *Movilidad laboral e informalidad: La situación argentina en el período 1998–2006.* CEDLAS.

Harris, J. R. y Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *American Economic Review, 60*(1), 126–142.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies, 11*(1), 61–89.

Heckman, J. y Sedlacek, G. (1985). Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy, 93*(6).

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900–1902.

International Labour Office. (1972). *Employment, incomes and equality: A strategy for increasing productive employment in Kenya.* ILO.

Jung, J. y Katz, R. (2026). *Impacto de la inteligencia artificial en el empleo y la informalidad en América Latina.* Comisión Económica para América Latina y el Caribe (CEPAL).

Kritz, E. (1986). Análisis del S.I.U. en América Latina. En L. Sáinz (Ed.), *CEDLA-FLACSO-ILDIS.*

Leibenstein, H. (1957). *Economic backwardness and economic growth.* John Wiley and Sons.

Lemieux, T. (2006). The "Mincer equation" thirty years after Schooling, Experience, and Earnings. En S. Grossbard (Ed.), *Jacob Mincer: A pioneer of modern labor economics* (pp. 127–145). Springer.

Levy, S. (2008). *Good intentions, bad outcomes: Social policy, informality, and economic growth in Mexico.* Brookings Institution Press.

Levy, S. (2018). *Under-rewarded efforts: The elusive quest for prosperity in Mexico.* Banco Interamericano de Desarrollo.

Levy, S. y Székely, M. (2016). Más escolaridad, ¿menos informalidad? Un análisis de cointegración para México. *El Trimestre Económico.*

Lewis, A. (1954). Economic development with unlimited supplies of labor. *The Manchester School.*

Longhi, A. (1998). *El trabajo y la economía informal. Discusión sobre enfoques teóricos.* Universidad de la República, Facultad de Ciencias Sociales.

Longhi, A. (2007). *El trabajo y la economía informal. Discusión sobre enfoques teóricos.* Facultad de Ciencias Sociales.

López Lapo, J. L. y Sarmiento Castillo, G. (2019). Determinantes de la brecha salarial en Ecuador: Análisis bajo un modelo minceriano relacionado con variables agregadas dummy. *Espirales Revista Multidisciplinaria de Investigación Científica, 3*(29).

Lubell, H. (1974). *Urban development and employment.* Geneva: International Labour Office.

Maia de Souza, C., Ponce-Lopez, R., Peraza-Mues, G. G., Dominguez-Cristerna, A. A., y Miller, E. J. (2025). Imputing informal workers for transportation modeling in Latin America by the use of machine learning techniques. *Journal of Transport Geography.*

Maloney, W. (2004). Informality revisited. *World Development, 32*(7).

Mincer, J. (1972). Time series changes in personal income inequality in the United States. *Journal of Political Economy, 80*(3).

Mincer, J. (1974). *Schooling, experience, and earnings.* National Bureau of Economic Research.

Moser, C. (1978). *Informal sector or petty commodity production: Dualism or independence in urban development.* Pergamon Press.

Mullainathan, S. y Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review, 14*(3), 693–709.

Perry, G. y Maloney, W. (2007). *Informalidad: Escape y exclusión.* Banco Mundial.

Piore, M. (1983). Labor market segmentation: To what paradigm does it belong? *American Economic Review, 73*(2).

Portes, A., Castells, M. y Benton, L. (1989). *The informal economy: Studies in advanced and less developed countries.* The Johns Hopkins University Press.

Portes, A. (2004). *La economía informal.* CEPAL.

Portes, A. y Schauffler, R. (1993). Competing perspectives on the Latin American informal sector. *Population and Development Review, 19*(1).

Prebisch, R. (1949). Introduction: The economic development of Latin America and its principal problems. *Economic Survey of Latin America 1948.* CEPAL.

Ribeiro, M. T., Singh, S. y Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Robles, S., Ponce, P., Alvarado, R., y Ortiz, C. (2019). Efecto del capital humano en la brecha de ingresos: Un enfoque utilizando propensity score matching. *Revista Economía y Política*, 25–47.

Rostow, W. (1960). *The stages of economic growth: A non-communistic manifesto.* Cambridge University Press.

Tokman, V. E. (1978). Las relaciones entre los sectores formal e informal. *Revista de la CEPAL.*

Tokman, V. E. (1989). Policies for a heterogeneous informal sector in Latin America. *World Development, 17*(7), 1067–1076.

Tokman, V. E. (2001). De la informalidad a la modernidad. *Revista Economía (PUCP)*, (48), 153–178.

Uribe, J. I. (2004). *Una propuesta de conceptualización y medición del sector informal.* Universidad del Valle.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–28.

Anexos:

Anexo I:

Cuaderno de Trabajo (Jupyther Notebook), adjunto a este documento.

Anexo II:

Ampliación del estudio histórico de la informalidad:

Aunque el término "sector informal" fue introducido formalmente por la OIT en su informe sobre Kenia en 1972, sus raíces conceptuales se remontan a debates académicos de las décadas de 1950 y 1960, en un contexto de reconstrucción económica y redefinición de paradigmas sociales tras la posguerra (OIT, 1972).

Las discusiones sobre el rumbo que debía tomar el desarrollo y los retos que planteaban las nuevas formas de empleo surgidas en la posguerra también alcanzaron a las economías en proceso de industrialización. En ese contexto, aportes como los de Raúl Prebisch con su planteamiento de Centro y Periferia (1949), Arthur Lewis con su teoría de la oferta ilimitada de trabajo (1954), Harvey Leibenstein con el Big Push (1957) y Rostow con sus etapas del crecimiento económico (1960) reflejaban la necesidad de comprender el papel de los países en un escenario de transformación acelerada. Para las naciones emergentes, definir su estrategia de desarrollo se convirtió en una prioridad fundamental para los Estados.

En ese escenario, la producción académica se concebía como un medio viable para orientar el tránsito de las economías hacia un crecimiento sostenido. No obstante, más allá de la solidez técnica de las propuestas, las ideas solían verse condicionadas por el contexto geográfico y político de donde surgían. En un mundo marcado por la polarización, ninguna teoría logró trascender plenamente sus fronteras de origen ni contar con el respaldo político necesario para aplicarse de manera universal (Bangasser, 2009).

Durante este periodo, en el que los países en desarrollo intentaban definir su rumbo y su modelo de crecimiento, el papel de las Naciones Unidas y de sus agencias especializadas, como la Organización Internacional del Trabajo, resultó decisivo. Al no estar directamente subordinadas a los Estados, estas instituciones podían mantener cierta independencia, lo que les otorgaba un margen de protección frente a las tensiones ideológicas, los prejuicios y los estereotipos propios de la Guerra Fría.

La posición estratégica desde la que la OIT difundía sus propuestas le otorgaba la ventaja de alcanzar una amplia visibilidad y objetividad, además de estar relativamente libre de los sesgos ideológicos asociados al patrocinio estatal o regional de las economías hegemónicas. En este contexto, amparada por el marco institucional de las Naciones Unidas, la OIT desempeñó un papel clave en la planificación de los Estados, centrándose en un ámbito decisivo: el mercado laboral. Para ello, impulsó la creación de estructuras institucionales que le permitieran abordar este tema con rigor, especialmente en los países en desarrollo. En paralelo, el Programa de las Naciones Unidas para el Desarrollo (PNUD) destinó recursos a la conformación del Departamento para el Desarrollo de Recursos Humanos, con el propósito de apoyar a los Estados en la organización y planificación de su mercado de trabajo, sin importar su orientación política. Una vez consolidada esta infraestructura, los primeros hallazgos sobre el estudio del sistema laboral de las economías emergentes generaron inicialmente confusión y posteriormente desconcierto:

Según sostienen Perry & Maloney (2007), incluso con el cumplimiento de condiciones macroeconómicas alineadas con supuestos ambiciosos de crecimiento, productividad e inversión, la cantidad de empleos generados resultaba insuficiente frente al tamaño de la fuerza laboral en las economías en desarrollo. En consecuencia, un número considerable de personas aspiraba a insertarse en los llamados "empleos modernos", pero no lograba ser absorbido por la demanda existente. Paralelamente, parte de la población económicamente activa quedaba fuera de los puestos planificados, aunque continuaba desempeñando actividades productivas, lo que implicaba que estaba ocupada en términos económicos. Este segmento, que operaba al margen de los empleos formales previstos, fue identificado como el sector informal. Tal caracterización coincide con lo señalado por la OIT (1972), que introdujo el concepto para describir a quienes generaban ingresos provenientes de actividades no reguladas ni reconocidas por el marco institucional.

En este tipo de actividades, a los ojos de la OIT se enmarcan los "sirvientes mal pagados", artesanos, granjeros, entre otros. Sin embargo, para la mayoría de los Estados, como incluso para la misma OIT, este sería un "problema transitorio", que tendría su fin como consecuencia del desarrollo de las economías, cuando el empleo formalmente planeado absorba a quienes desempeñan actividades informales; de modo que no se necesitaba más que monitorear el problema, caracterizarlo y describirlo continuamente; con eso bastaba. Cabe destacar que el empleo formalmente planeado hace referencia a la denominación empleada por la OIT para catalogar las plazas de empleo programadas por el Estado, el sector privado y actores institucionalmente reconocidos.

Con el paso del tiempo, los estudios de la OIT evidenciaron que la insuficiencia de empleos no era un fenómeno coyuntural, sino una brecha estructural del mercado laboral que requería un análisis profundo de sus causas y dinámicas. En consecuencia, la organización recomendó que los programas de desarrollo económico no trataran al empleo como un resultado secundario del crecimiento de la inversión o de la productividad, sino como un objetivo central que demandaba políticas específicas. Estas recomendaciones se sustentaban en la experiencia de países en desarrollo, donde se comprobó que la evolución del empleo no podía considerarse una variable de ajuste, sino que exigía intervenciones explícitas de la política pública (Bangasser, 2009). Hacia finales de la década de 1960, el estudio del mercado laboral y, en particular, la comprensión de la informalidad, empezaron a consolidarse como temas centrales en el debate académico internacional, cuestionando los supuestos de la economía clásica sobre salarios de equilibrio y el vaciamiento de los mercados.

Así, según Enríquez (2014), desde 1972, la OIT destinó recursos significativos para profundizar en el estudio de la informalidad, organizando las llamadas "misiones comprensivas". Estas iniciativas buscaban identificar las particularidades y los patrones comunes del fenómeno en distintas economías en desarrollo, con el fin de enriquecer la comprensión del mismo y generar evidencia comparativa. Los equipos estaban conformados por especialistas de diversas disciplinas, quienes se enfocaban en analizar de manera rigurosa el mercado laboral urbano informal de cada país. Entre estas experiencias destaca la misión liderada por Keith Hart, cuyo trabajo en Ghana aportó elementos fundamentales para conceptualizar el sector informal como categoría analítica dentro de la economía del desarrollo (Hart, 1973; OIT, 1972).

El estudio realizado por Keith Hart en Ghana a inicios de la década de 1970 resultó decisivo al cuestionar la visión negativa que hasta entonces predominaba sobre la informalidad. Antes de su aporte, el enfoque dominante asociaba este fenómeno con ocupaciones marginales, como las de vendedores ambulantes o trabajadores subempleados en las grandes ciudades. Hart demostró que, lejos de ser únicamente un reflejo de precariedad, la informalidad podía constituir una estrategia económica eficiente, capaz de generar ingresos y sostener a quienes la integraban, aun cuando las unidades productivas fueran

pequeñas, con escaso capital y tecnología limitada (Hart, 1973). En este sentido, el problema no residía en la existencia de la informalidad, sino en su interpretación reduccionista. Tal como señalan Tokman (1978) y la OIT (1972), la falta de comprensión de su dinámica llevó a que las actividades informales fueran ignoradas o desincentivadas por los gobiernos, pese a que representaban una alternativa real frente al desempleo.

Según Enríquez (2014), junto con las misiones comprensivas, el Programa Mundial de Empleo de 1974 marcó un punto de inflexión en el análisis de la informalidad. Entre sus principales referentes se encuentra Harold Lubell, quien, en su obra sobre desempleo urbano en economías emergentes (1971), simplificaron categorías conceptuales que dificultaban el estudio del mercado laboral, como las nociones ambiguas de desempleo encubierto, subempleo visible o subempleo oculto. Estos aportes permitieron avanzar hacia una comprensión más clara de las dinámicas laborales en los países en desarrollo.

En el ámbito de la informalidad, sus primeros estudios se concentraron únicamente en el mercado laboral urbano, lo que generó importantes limitaciones analíticas. Quedaba sin respuesta la pregunta sobre la informalidad rural: ¿era marginal o irrelevante, o presentaba dinámicas propias que merecían atención? Al no considerar este segmento, se asumió que la experiencia urbana bastaba para inferir conclusiones aplicables al conjunto de la informalidad. Como resultado, el fenómeno fue interpretado como un "sumidero laboral", en el que se ubicaban tanto quienes no lograban insertarse en el sector formal como los migrantes rurales que no podían acceder a empleos modernos (Longhi, 2007). En ese contexto, la informalidad se consolidó como uno de los temas más discutidos a nivel global en la década de 1970, percibida como un sector de último recurso y una alternativa poco deseada, sin reconocer plenamente las funciones económicas que podía desempeñar (Tokman, 1978; OIT, 1972).

Tras estos esfuerzos, el estudio de la informalidad adquirió gran relevancia y la OIT se convirtió en el eje del debate, lo que generó aportes que, en ocasiones, se contradecían entre sí. A este proceso académico se sumaron factores políticos y tensiones externas que limitaron la objetividad de los análisis, fragmentaron aún más el enfoque y desalentaron nuevas investigaciones. Un hecho decisivo fue la salida de Estados Unidos de la OIT en 1973, lo que redujo significativamente el presupuesto de la organización y puso de manifiesto que buena parte de la producción académica respondía a presiones políticas (Bangasser, 2009). En palabras del entonces director del Programa Mundial de Empleo, se advertía de que la continuidad de los estudios sobre el sector informal no debía justificarse únicamente por la tradición investigativa, sino también por la necesidad real de los constituyentes. Esta postura reforzó la llamada "visión miseralista", influida por los trabajos de Paul Bairoch y Harold Lubell, quienes interpretaban la informalidad como un fenómeno indeseado y residual. A partir de allí se consolidó una hoja de ruta marcada por un enfoque paternalista, que tendía a desconocer las particularidades nacionales y a reducir la informalidad a un problema de atraso estructural (Lubell, 1974; Bairoch, 1971; Tokman, 1978).

Según Enríquez (2014), este episodio puso de manifiesto que la objetividad de organismos internacionales como la OIT no podía garantizarse únicamente por su carácter independiente respecto de los Estados. En la práctica, aun cuando sus programas y objetivos se presentaban como respuestas técnicas a las necesidades nacionales, existía un amplio margen para la influencia política, la dependencia de modelos de desarrollo dominantes y las presiones provenientes de las fuentes de financiamiento. Asimismo, las dinámicas burocráticas internas y los compromisos informales podían afectar la orientación de sus iniciativas, generando implicaciones sociales significativas e incluso favoreciendo intereses de carrera personales (Bangasser, 2009). En este sentido, diversos autores han señalado que la institucionalidad de la OIT, aunque pionera en el estudio de la informalidad, estuvo condicionada por tensiones políticas y económicas propias del contexto internacional de la época.

Independientemente del giro que la OIT le dio al estudio de la informalidad, y más allá de los matices, contrastes e incluso contradicciones de sus investigaciones, no se puede negar que su labor resultó determinante para alimentar el cuestionamiento e indagación sobre el tema, pues su estructura organizacional le permitió presentar sus argumentos y difundirlos con relativa facilidad, y quizá este es uno de sus mayores méritos.

Fruto de la labor de la OIT durante la década de los 70, el estudio de la informalidad ya había tomado cuerpo, pero también invitaba a cuestionar la validez de sus argumentos. Así, para 1980, eran múltiples los cuestionamientos, la refutación y la confrontación en espacios de debate, entre las distintas escuelas de pensamiento económico, que habían empezado a preocuparse con rigurosidad académica específicamente por el tema de la informalidad, ajustando sus teorías a esta aplicación en particular. Entre los aportes más importantes surgidos en la década de los 80 se puede distinguir a:

Más allá de los matices, los contrastes e incluso las contradicciones presentes en sus investigaciones, la labor de la OIT fue decisiva para impulsar el debate sobre la informalidad. Su estructura institucional le permitió difundir argumentos con relativa facilidad, lo que constituye uno de sus principales aportes al posicionamiento del tema en la agenda académica internacional.

Gracias al trabajo desarrollado durante la década de 1970, el estudio de la informalidad adquirió consistencia, aunque también suscitó cuestionamientos sobre la validez de sus planteamientos. Para 1980, las discusiones se intensificaron en diversos espacios académicos, donde las escuelas de pensamiento económico comenzaron a ajustar sus teorías para abordar específicamente la informalidad. En este contexto, surgieron contribuciones relevantes que marcaron la década y permitieron reinterpretar la informalidad como un fenómeno de exclusión y de salida. Entre los aportes más importantes surgidos en la década de los 80 se puede distinguir a:

El pensamiento neomarxista (1981): Esta escuela de pensamiento adquirió gran relevancia en el análisis de la informalidad. Su planteamiento central señalaba la contradicción que supone la existencia del sector informal dentro del modelo de acumulación capitalista, interpretada como una "inconsistencia necesaria". Mientras que el capitalismo tiene como finalidad última la acumulación de capital, el sector informal urbano (SIU) se orienta principalmente a la subsistencia de la fuerza laboral que lo integra, priorizando el consumo inmediato sobre la acumulación. En este sentido, la informalidad cumple una función estructural: mantener bajos los salarios en las economías en desarrollo, lo que refuerza la lógica de acumulación del capital formal (Longhi, 1998; Portes & Castells, 1989). Este enfoque permitió comprender la informalidad no como un simple residuo del sistema, sino como un componente funcional de la dinámica del capitalismo periférico.

Heckman y Sedlacek (1985): Heckman y Sedlacek (1985), reconocidos como uno de los principales exponentes de la Escuela de Chicago en el análisis de la informalidad, planteó que este sector debía entenderse como un "bien normal" en la función de producción. Desde su perspectiva, la baja productividad explica en gran medida por qué los trabajadores permanecen en la informalidad. Factores como el nivel educativo limitado, la escasa especialización y las habilidades poco tecnificadas determinan que este segmento se concentre en actividades que no requieren mano de obra altamente calificada. De ahí que la informalidad tenga una presencia tan amplia en América Latina, pues se caracteriza por requerir una mínima inversión de capital y absorber abundante fuerza laboral de baja calificación, pero dispuesta a emplearse en condiciones precarias (Heckman & Sedlacek, 1985; Maloney, 2004).

Ernesto Kritz (1986): En la segunda mitad de la década de 1980, Ernesto Kritz (1986) planteó que el estudio de la informalidad se encontraba atrapado en dos vertientes heredadas de las escuelas marxista

y neoclásica. Para él, esta situación representaba una limitación importante, ya que ambos enfoques tendían a utilizar la informalidad como argumento accesorio para validar sus teorías principales, sin atender directamente a las particularidades del fenómeno. En consecuencia, ninguno de los dos logró superar la visión dualista que separaba rígidamente el sector formal del informal, lo que impedía analizar las interacciones entre ambos y reducía la capacidad del debate para responder a las necesidades de las poblaciones afectadas.

El aporte de Kritz consistió en proponer un enfoque que reconociera la existencia de la informalidad dentro de la formalidad, una dimensión hasta entonces poco explorada. Su perspectiva buscaba superar el dualismo, enfatizando la articulación y la simbiosis entre ambos sectores. De este modo, resaltó el carácter heterogéneo de la informalidad, entendida no como un único modo de producción, sino como un conjunto diverso de formas organizativas que no podían homogeneizarse ni simplificarse. Este planteamiento abrió la puerta a estudios posteriores que reconocieron la multiplicidad de dinámicas informales y su interacción con la economía formal (Kritz, 1986; Tokman, 1989; Portes & Castells, 1989).

Wim Dierckxsens (1989): El trabajo de Wim Dierckxsens (1989) representó un giro importante en el análisis de la informalidad en las economías en desarrollo. Su propuesta introdujo un matiz estructural novedoso al señalar que la informalidad emerge de la interacción entre dos condiciones: la existencia de economías en transición y la presencia de una fuerza laboral fácilmente sustituible. El primer factor se relaciona con una sobrepoblación vulnerable que debe desenvolverse en mercados altamente concentrados y dominados por capital extranjero; el segundo, con la baja tecnificación de los trabajadores, que los vuelve reemplazables en el proceso productivo. En este contexto, la edad se convierte en un indicador clave de productividad: según su estudio, los trabajadores mayores de 35 años tienden a ser desplazados de los empleos formales y a encontrar en el sector informal una alternativa de subsistencia. Con este planteamiento, Dierckxsens aportó un enfoque estructural distinto, que enriqueció el debate académico al mostrar cómo la informalidad debía entenderse como parte integral de las dinámicas de transición económica en países periféricos.

Las investigaciones destacadas de la década de 1980 aportaron múltiples perspectivas para observar y analizar la informalidad. Sin embargo, estos esfuerzos no lograron consolidar una agenda de investigación definida ni establecer un horizonte claro para el desarrollo del campo. Persistían numerosas preguntas sin respuesta y aspectos pendientes por abordar. En consecuencia, hacia finales de esa etapa, el estudio de la informalidad apenas comenzaba a estructurarse como un área de interés académico, pero aún carecía de la capacidad de ofrecer soluciones efectivas a las necesidades de las sociedades latinoamericanas, que requerían propuestas más rigurosas y sostenidas en el tiempo.

Durante la década de 1990, el estudio de la informalidad adquirió un compromiso más sostenido y de largo plazo. Hernando de Soto identificó que se trataba de una demanda urgente y propuso un enfoque académico integral para su análisis. Su aporte consistió en cuestionar la visión negativa y paternalista con la que tradicionalmente se asociaba la informalidad, señalando que este sesgo condicionaba las conclusiones de los investigadores. En lugar de ello, planteó la necesidad de adoptar una perspectiva económica basada en costos, beneficios e incentivos individuales, lo que permitía una lectura más objetiva del fenómeno (De Soto, 1989).

Desde esta óptica, las actividades informales se interpretaban como expresión de un dinamismo empresarial popular, resultado de las rigideces del Estado. Según De Soto (1989), los Estados se concentraban en regular y controlar, sin diseñar mecanismos eficientes, lo que convertía la formalidad en un privilegio reservado a una élite capaz de asumir los costos legales. Así, frente a la concepción de

la OIT y el PREALC, que entendían la informalidad como un mecanismo de supervivencia ante la insuficiente generación de empleo moderno, De Soto la definió como la irrupción de fuerzas reales de mercado en economías constreñidas por la regulación estatal (Portes & Schauffler, 1993).

El debate académico de los noventa se enriqueció con este nuevo enfoque, aunque también se polarizó. Por un lado, persistía la línea de investigación impulsada por la OIT; por otro, emergía la propuesta de De Soto, que pese a su novedad, se mostraba sólida y estructurada. En consecuencia, los argumentos se situaron en extremos que difícilmente podían conciliarse, lo que reflejaba la tensión entre visiones institucionalistas y perspectivas de mercado en el estudio de la informalidad.

Durante la primera década del siglo XXI, el estudio de la informalidad en América Latina se consolidó como un campo de investigación interdisciplinario, en el que convergieron enfoques económicos, sociológicos y de políticas públicas, pero de manera limitada los de las ciencias exactas o computacionales. Uno de los aportes más influyentes provino de William Maloney (2004)**,** quien cuestionó la visión tradicional de la informalidad como un espacio exclusivamente de exclusión y precariedad. Su propuesta introdujo la noción de la informalidad como una "salida" voluntaria**,** en la que ciertos trabajadores y microempresarios optan por permanecer fuera de la regulación estatal debido a los altos costos de la formalización y la rigidez institucional. Este enfoque permitió comprender la informalidad como un fenómeno heterogéneo, en el que coexisten dinámicas de exclusión estructural y estrategias de emprendimiento popular. La década, además, estuvo marcada por estudios de organismos multilaterales como el Banco Mundial y la CEPAL, que comenzaron a vincular la informalidad con problemas de productividad, competitividad y sostenibilidad fiscal, ampliando el debate más allá del mercado laboral hacia la estructura económica en su conjunto.

En los últimos años, el análisis de la informalidad ha evolucionado hacia enfoques más estructurales y multidimensionales, incorporando metodologías cuantitativas avanzadas y perspectivas institucionales. Destaca el trabajo de Santiago Levy (2018)**,** quien sostiene que la informalidad en América Latina no es un fenómeno transitorio, sino un rasgo persistente vinculado a sistemas fiscales fragmentados y a regímenes de protección social ineficientes. Levy sostiene que la informalidad perpetúa la baja productividad y limita el crecimiento económico, al tiempo que reproduce desigualdades en el acceso a derechos laborales y sociales. Paralelamente, la CEPAL (2020, 2024) ha enfatizado la necesidad de analizar la informalidad en relación con nuevas dinámicas como la digitalización, el trabajo en plataformas, la precarización juvenil y las brechas de género. Estos estudios muestran que la informalidad no desaparece con la modernización tecnológica, sino que se adapta y persiste en nuevas formas, e incluso puede potenciarse en contextos de disrupción tecnológica, de aparición de nuevas inteligencias y del afloramiento de la industria 4.0, que exige marcos regulatorios innovadores y políticas públicas integrales.

Informalidad y la Industria 4.0 (2020–2025): Los informes recientes de la CEPAL (2020, 2024) han marcado un giro en la comprensión de la informalidad frente a los desafíos de la digitalización y la inteligencia artificial. Levy sostiene que la informalidad en América Latina constituye un rasgo estructural del subdesarrollo, vinculado a sistemas fiscales fragmentados y regímenes de protección social débiles. En el marco de la Industria 4.0, esta condición se vuelve más compleja: la incorporación de tecnologías digitales e inteligencia artificial tiende a beneficiar principalmente a los sectores formales y altamente calificados, lo que amplía la brecha con los trabajadores informales. La CEPAL advierte que la informalidad no desaparece con la digitalización, sino que se transforma en nuevas modalidades, como el trabajo en plataformas digitales, los microemprendimientos virtuales y los servicios mediados por aplicaciones, que reproducen la precariedad laboral en formas híbridas. En este escenario, los trabajadores informales enfrentan una doble exclusión: por un lado, la tradicional, derivada de la falta

de acceso a la seguridad social y a los derechos laborales; y por otro, la tecnológica, vinculada a la carencia de habilidades digitales y al limitado acceso a la infraestructura. Con este planteamiento, los estudios recientes aportan un matiz estructural novedoso: la informalidad debe entenderse como un fenómeno que se adapta y persiste en la era digital, lo que revela la necesidad de políticas públicas que integren la capacitación tecnológica, la regulación de nuevas formas de empleo y la ampliación de la protección social para evitar que la Industria 4.0 profundice las desigualdades en economías periféricas como la ecuatoriana.

Juan Jung y Raúl Katz (CEPAL, 2026): El aporte de Jung y Katz (2026) constituye un esfuerzo relevante para situar la informalidad en el marco de la transformación digital y de la adopción de la inteligencia artificial en América Latina. Sus hallazgos muestran que, si bien la IA puede incrementar la productividad del trabajo calificado y abrir oportunidades de innovación, también corre el riesgo de profundizar las brechas estructurales entre trabajadores formales e informales, pues existe un nicho de precarización asociado a la nueva tecnologización de los servicios. En este sentido, la informalidad se convierte en un espacio de vulnerabilidad frente a la digitalización, ya que los trabajadores informales carecen de acceso a capacitación especializada y a mecanismos de financiamiento que les permitan integrarse en la nueva economía digital, por un lado; y por otro, una vez insertos en la economía digital permanecen vulnerables, sin estabilidad laboral y operando en la sombra de costos laborales artificialmente competitivos. Con este planteamiento, Jung y Katz aportan un matiz estructural novedoso: la informalidad no solo refleja desigualdades históricas, sino que también se proyecta como un factor de exclusión tecnológica en la era de la Industria 4.0.

Foro Económico Mundial (2025): El informe del Foro Económico Mundial (2025) subraya que en el Sur Global, y particularmente en América Latina, más del 60% de la fuerza laboral permanece en la informalidad, lo que convierte a la región en un caso paradigmático de cómo la digitalización interactúa con mercados laborales fragmentados. El estudio revela que los trabajadores informales ya utilizan herramientas digitales como WhatsApp, pagos móviles y sistemas de reputación en línea para sostener sus actividades económicas, pero lo hacen en ausencia de marcos regulatorios adecuados. Las precarias reglas entre las que se desenvuelven contemplan la administración de comisiones y márgenes de ganancia, así como el reconocimiento de costos fijos sujetos a la venta o a la reservación de un bien o servicio. De modo que sus derechos son incipientes y fenecen al concluir la demanda del producto o servicio. Esto abre un amplio escenario de desprotección e inestabilidad. El hallazgo central es que la informalidad no está disminuyendo con la digitalización, sino que está evolucionando hacia formas digitales, reproduciendo la precariedad en nuevas modalidades. Este diagnóstico plantea un desafío crucial: diseñar políticas públicas que reconozcan estas prácticas digitales e integren estas prácticas en sistemas de protección social y laboral, evitando que la digitalización se convierta en un mecanismo de precarización ampliada.

Marisol Argueta y Cathy Li (Foro Económico Mundial, 2025): En el informe *América Latina en la Era Inteligente*, Argueta y Li (2025) enfatizan que la inteligencia artificial puede convertirse en un motor de competitividad regional, pero advierten que si las economías inteligentes no sitúan a las personas en el centro de sus estrategias, la informalidad se transformará en un espacio de exclusión tecnológica. Según su análisis, millones de trabajadores podrían quedar fuera de los beneficios de la digitalización, lo que generaría una nueva forma de desigualdad estructural. La informalidad, en este contexto, no solo refleja carencias institucionales y fiscales, sino también la incapacidad de los Estados para garantizar una transición inclusiva hacia la economía digital. Su planteamiento aporta un matiz relevante y pertinente: la informalidad debe entenderse como un fenómeno que se reproduce y se adapta en la era de la IA, y cuyo abordaje requiere políticas integrales que combinen innovación tecnológica con inclusión social.

Anexo III:

Análisis de Sensibilidad del Tratamiento de Valores Atípicos en el Ingreso Laboral

El presente análisis responde a la necesidad de evidenciar el efecto de la aplicación de la técnica de winsorización al 2% aplicada al ingreso laboral (ingrl) en la etapa de preprocesamiento, sobre la estabilidad del coeficiente de informalidad en la ecuación de Mincer extendida. Para ello, se presenta primero el resultado comparativo entre tres tratamientos del ingreso:

El punto de partida: la winsorización no actúa sobre datos crudos

La winsorización es, en efecto, una técnica estandarizada dentro del método KDD (Knowledge Discovery in Databases) para el tratamiento de valores atípicos, y su inclusión en el preprocesamiento de este trabajo responde a esa práctica reconocida. Sin embargo, antes de aplicarla, el ingreso laboral ya había sido sometido a dos depuraciones previas, ambos estándares en el manejo de encuestas de hogares e ingresos laborales en la región:

En primer lugar, los ingresos declarados bajo el código 999999 se recodifican como valores perdidos (NaN). Es importante precisar que este código no identifica un valor atípico en el sentido estadístico del término, sino una categoría administrativa de la encuesta que significa "no aplica" o "no especifica": conceptualmente, no es un ingreso extremo, sino la ausencia de un ingreso reportado. En segundo lugar, la muestra de estimación se restringe a ingresos laborales positivos (ingrl > 0), como parte de la definición misma de la población objetivo -trabajadores ocupados con ingreso laboral-, no como un criterio adicional de depuración de atípicos.

Estas dos depuraciones, aplicadas de forma previa y estándar, ya delimitan al ingreso laboral dentro de un rango sustancialmente más acotado que el de la base completa de la ENEMDU, que incluye a personas fuera de la fuerza laboral, desempleadas, o sin ingreso declarado. En consecuencia, es esperable que buena parte de lo que la winsorización busca corregir (valores extremos o inválidos) ya haya sido removido antes de que esta técnica se aplique, dejando un margen reducido para que opere sobre la muestra de interés. Pese a ello, se incorpora esta técnica como un paso previo al modelamiento, con el objetivo de dar cumplimiento a una de las fases estandarizadas del método KDD.

Análisis comparativo

Bajo esa consideración, se estimó la ecuación de Mincer extendida sobre tres versiones del ingreso laboral, manteniendo las dos depuraciones previas (recodificación de 999999 y restricción a ingresos positivos) constantes en los tres casos, y variando únicamente el tratamiento adicional aplicado sobre las colas de la distribución:

Cuadro. Sensibilidad del coeficiente de es_informal (Mincer extendida) al tratamiento de valores atípicos.

Tratamiento	N	Coef. es_informal	Error est.	Penalización (%)	R^2
Winsorización 2% (especificación utilizada)	77.785	-0,3670	0,0065	30,7%	0,4927
Sin tratamiento adicional de colas	77.785	-0,3670	0,0065	30,7%	0,4927

Truncamiento 2% (elimina, no acota, las colas)	76.360	-0,3557	0,0059	29,9%	0,4939

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Re plicabilidad: notebook adjunto, sección 6.1.2 (complemento).

Explicación sobre la coincidencia entre la winsorización global y el ingreso sin tratamiento

El resultado más llamativo del cuadro anterior es que el coeficiente de es_informal bajo winsorización al 2% y bajo el ingreso sin ningún tratamiento adicional de colas es idéntico (β = -0,3670 en ambos casos, con el mismo error estándar y el mismo R^2). Esta coincidencia no es fortuita, y su explicación conecta directamente con el argumento planteado al inicio de este documento: la winsorización al 2% se calculó sobre la totalidad de la base ENEMDU (334.786 personas), una población que incluye a quienes están fuera de la fuerza laboral, desempleados, o sin ingreso declarado. Los percentiles 2 y 98 de esa base completa (0 y 2.187 dólares) resultan, por esa razón, sustancialmente más anchos que el rango de ingresos efectivamente presente dentro de la muestra utilizada para estimar la ecuación de Mincer extendida (111.122 personas), la cual se obtiene aplicando dos filtros en cadena sobre la base completa: primero, la restricción a población ocupada de 15 años o más con ingreso laboral positivo (137.588 personas); y segundo, dentro de ese grupo, la restricción adicional a quienes cuentan con datos completos en todas las variables que emplea la ecuación extendida -escolaridad, experiencia, sexo, área, estado civil, tamaño de empresa, ocupación, rama de actividad y quintil de ingreso del hogar-, lo que descarta 26.466 personas adicionales con algún dato faltante y deja la muestra final en 111.122 personas, dentro de la cual el ingreso oscila entre 1 y 1.908 dólares. En otras palabras, para cuando la winsorización se aplica, las dos depuraciones previas descritas al inicio de este documento (la recodificación de los códigos 999999 y la restricción a ingresos laborales positivos) ya han retirado de la muestra tanto los códigos de no respuesta como las observaciones sin ingreso laboral positivo, dejando un rango de valores que la winsorización, calculada sobre una población mucho más amplia y heterogénea, ya no tiene margen para modificar. El ingreso "sin tratamiento" de este ejercicio no es, en sentido estricto, ingreso sin depurar: ya incorpora esas dos depuraciones previas, que son precisamente las que absorben la mayor parte de lo que podría considerarse un valor atípico o inválido en esta base de datos. La winsorización, aplicada después, actúa entonces sobre un remanente en el que prácticamente no quedan observaciones capaces de superar sus propios umbrales.

Sobre la validez de winsorizar la distribución completa del ingreso respecto a la utilización de datos truncados localmente al 2%

Aclarado lo anterior, para los fines de este ejercicio se considera adecuada y válida la aplicación de la winsorización como una fase genérica y estandarizada del tratamiento del ingreso laboral sobre toda su distribución, y no exclusivamente sobre la muestra específica de estimación. Aplicar la winsorización sobre la base completa, en lugar de recalcularla para cada submuestra de trabajo, permite garantizar la consistencia global del tratamiento del ingreso a lo largo de todo el estudio, evitando que el mismo valor de ingreso reciba un tratamiento distinto según el modelo o la sección del trabajo en que se utilice.

Sin embargo, con el fin de evidenciar el efecto de trabajar con datos truncados frente a la winsorización adoptada, el Cuadro anterior presenta también los resultados de la ecuación de Mincer extendida bajo un truncamiento del 2%, calculado una vez descontados los códigos de "no aplica" o "no especifica" (999999) y los valores de ingreso iguales o menores a 0. Es decir, el 1% inferior y el 1% superior que se eliminaron bajo este tratamiento corresponden a ingresos laborales genuinamente reportados y

extremos dentro de la población ya depurada por esos dos filtros (por debajo de 16 dólares o por encima de 1.500 dólares mensuales), y no a códigos de no respuesta ni a observaciones fuera de la población objetivo, que ya habían sido excluidos en una etapa anterior e independiente del truncamiento.

Bajo este tratamiento sí se evidencia un efecto, tanto en el coeficiente de es_informal (de -0,3670 bajo winsorización a -0,3557 bajo truncamiento) como en el R^2 (de 0,4927 a 0,4939) y, de forma más marcada, en el tamaño de la muestra (de 77.785 a 76.360 observaciones en el conjunto de entrenamiento). Ese efecto, sin embargo, es consecuencia directa de la eliminación de observaciones (1.425 personas menos en el conjunto de entrenamiento), lo que se considera desfavorable en términos de calidad y aprovechamiento de la información disponible. Es precisamente por esta razón que este trabajo emplea los resultados derivados de la winsorización, y no del truncamiento, como tratamiento definitivo del ingreso laboral: la winsorización acota los valores extremos sin sacrificar observaciones, mientras que el truncamiento, pese a operar sobre el mismo tipo de ingresos genuinos, solo logra su efecto a costa de reducir la muestra disponible para la estimación.

Conclusión respecto a la depuración de valores atípicos el ingreso laboral

En conjunto, este análisis de permite responder a la observación planteada con una explicación metodológica precisa, y no solo con una comprobación empírica de robustez: el umbral de winsorización elegido no altera los resultados de la ecuación de Mincer extendida porque las depuraciones previas del ingreso laboral (estándares en el manejo de encuestas de hogares en la región) ya habían delimitado la muestra de estimación a un rango dentro del cual esa winsorización, calculada sobre la población completa, no tiene margen de acción adicional. Y allí donde sí existe un tratamiento con capacidad real de modificar la muestra mediante el truncamiento directo de ingresos extremos, en realidad su efecto sobre el coeficiente de interés es marginal (0,8 puntos porcentuales) y sus implicaciones en términos de la calidad de información, al deliberadamente dejar de contar con 1.425 registros, es significativa. Por ello se opta por la winsorización global a toda la distribución en lugar de por el truncamiento local. Las conclusiones que se derivan en consecuencia, indican que la penalización salarial de la informalidad es estable frente a las decisiones de preprocesamiento evaluadas en este ejercicio.

Justificación del umbral de winsorización al 2%

El umbral se deriva de forma directa de la construcción de la distribución del ingreso laboral en 100 percentiles: al dividir la distribución en cien partes iguales, el percentil 1 marca el límite por debajo del cual se ubica el 1% de los ingresos más bajos, y el percentil 99 marca el límite por encima del cual se ubica el 1% de los ingresos más altos. Aplicar winsorización a ambos extremos, acotando los valores por debajo del percentil 1 y por encima del percentil 99 al valor del propio percentil correspondiente equivale, en conjunto, a tratar el 2% de la distribución total: es decir, 1% en cada cola. Esta construcción, basada directamente en percentiles, es la que sustenta la elección de un umbral del 2% y no otro: representa el tratamiento del segmento más extremo de la distribución en cada dirección, sin extenderse a un segundo o tercer percentil que comenzaría a afectar observaciones menos atípicas y más representativas de la variabilidad genuina del ingreso.

Esta elección es, además, consistente con la práctica descrita por Kokic y Smith (1998) para el tratamiento de valores atípicos en encuestas de hogar, y se apoya en el principio general planteado por Tukey (1962): acotar el efecto de los extremos sin sacrificar el tamaño de la muestra, a diferencia de un tratamiento por eliminación directa o truncamiento.

Anexo IV:

Justificación Técnica: Valor Incremental de los Modelos de Clasificación Frente a una Línea Base

Ante la necesidad de establecer una línea base (baseline) más simple que la regresión logística y el Gradient Boosting Classifier, con el fin de cuantificar cuánto valor incremental real aporta cada modelo respecto a no realizar ningún ejercicio de modelamiento; se presenta a continuación las métricas de desempeño de los modelos de regresión logística y de gradient boosting classiffier frente a un modelo de mayorías. Esto permitirá determinar qué ganancia frente al azar proporcionan los modelos construidos, y con ello se dará mayor contexto de cuánto aporta el esfuerzo de modelamiento:

Para ello, se calculó el desempeño de un modelo de clasificación trivial basado en la clase mayoritaria que predice siempre la clase más frecuente de la muestra (informal, 54,3% del conjunto de prueba) sin utilizar ninguna variable explicativa y se contrastó contra los dos modelos ya estimados, evaluados sobre el mismo conjunto de prueba:

Cuadro. Desempeño comparado: modelo de mayoría, regresión logística y Gradient Boosting Classifier.

Modelo	Accuracy	Precisión	Recall	F1
Modelo de mayoría (línea base)	0,543	0,543	1,000	0,704
Regresión logística	0,820	0,817	0,861	0,838
Gradient Boosting Classifier	0,824	0,830	0,850	0,839

Fuente: elaboración propia a partir de ENEMDU-INEC (2025). Replicabilidad: cuaderno de trabajo adjunto, complemento posterior a la sección 6.2.8.

Dos de los indicadores habitualmente reportados no aportan evidencia útil al evaluar al clasificador de mayoría, por tratarse de resultados que se derivan matemáticamente de su definición y no del comportamiento de los datos. El AUC-ROC de un clasificador que predice siempre la misma clase es, por construcción, exactamente 0,5: al no generar ningún orden ni jerarquía entre observaciones, carece de la capacidad de discriminación que esta métrica mide, y su valor no depende de la muestra sobre la que se calcule. De forma análoga, el Recall del modelo de mayoría es, en este ejercicio, de 1,000: al predecir siempre la clase informal (la más frecuente), ningún caso informal queda sin identificar, pero ese resultado no refleja ninguna capacidad predictiva, sino una consecuencia mecánica de no distinguir nunca la clase contraria, o lo que es lo mismo, es únicamente el resultado de la definición del modelo y no una métrica válida de su desempeño real.

Por esta razón, el análisis centra la atención en las métricas de Precisión y F1, los dos indicadores de este ejercicio que sí varían en función de la calidad real de cada modelo y que, por tanto, permiten una comparación informativa entre el modelo de clasificación de mayorías (línea base) equivalente a lo que el azar reflejaría y el de los modelos efectivamente estimados.

La ganancia sustancial de ambos modelos sobre la línea base

Bajo estas consideraciones, la evidencia es clara: la regresión logística alcanza una Precisión de 0,817 y un F1 de 0,838, mientras que el Gradient Boosting Classifier alcanza una Precisión de 0,830 y un F1

de 0,839, frente a una Precisión de 0,543 y un F1 de 0,704 del modelo de mayoría. La ganancia de F1 de ambos modelos sobre la línea base (0,134 para la regresión logística y 0,136 para el Gradient Boosting) representa una mejora de más de trece puntos porcentuales sobre el punto de partida trivial, y confirma que ambos enfoques capturan una señal predictiva real y considerable a partir de las características de origen social, capital humano e inserción ocupacional del trabajador, y no una ganancia atribuible al azar o a la composición de la muestra.

El margen reducido entre modelos: un piso alto, no un techo bajo

La diferencia de F1 entre el Gradient Boosting Classifier y la regresión logística es de apenas 0,001 -incluso menor que la diferencia de 0,005 ya reportada en AUC-ROC en la sección 6.2.3-. Esta magnitud reducida no debe interpretarse como una limitación del modelo más sofisticado, sino como consecuencia directa de un acierto metodológico previo: la construcción cuidadosa de las variables explicativas (capital humano propio y heredado, tamaño del establecimiento, categoría ocupacional, rama de actividad y quintil de ingreso del hogar) ya captura, por sí sola, la mayor parte de la señal predictiva disponible en los datos. Cuando ambos modelos comparten prácticamente el mismo conjunto de insumos, el margen que resta para diferenciarse entre sí es, por definición, pequeño: mejorar sobre un piso ya alto es sistemáticamente más difícil que mejorar sobre un piso bajo, y ese piso alto es, en sí mismo, evidencia de la calidad del preprocesamiento de la información y de la construcción de variables realizado en las etapas previas de este trabajo.

Eficiencia computacional del Gradient Boosting Classifier

Conviene precisar, además, que la ganancia adicional del Gradient Boosting Classifier sobre la regresión logística no se obtuvo a costa de un uso intensivo de recursos computacionales. La selección de hiperparámetros de este modelo, descrita en la sección 6.2.2 del cuaderno de trabjo, se realizó mediante una comparación liviana entre un número acotado de configuraciones candidatas y no mediante una búsqueda exhaustiva de alto costo computacional; seleccionando la configuración con mejor desempeño de prueba entre ellas. La ganancia marginal documentada en este documento se logra, en consecuencia, sin incurrir en un sobrecosto de procesamiento desproporcionado respecto al beneficio obtenido, lo que constituye un segundo acierto metodológico complementario al de la construcción de variables.

Síntesis: el objetivo no es la disputa entre modelos, sino su efectividad conjunta

Finalmente, se debe destacar que el propósito de emplear dos modelos de clasificación en este trabajo nunca fue determinar cuál de los dos es superior entre sí, sino disponer de dos enfoques uno paramétrico e interpretable, otro no paramétrico y flexible, que permitieran identificar con solidez los perfiles de riesgo de informalidad. Bajo ese objetivo, la evidencia presentada permite determinar que tanto la regresión logística como el Gradient Boosting Classifier superan de forma amplia a un ejercicio de clasificación aleatorio o trivial, con ganancias de F1 superiores a trece puntos porcentuales sobre la línea base en ambos casos. Que la diferencia entre ambos modelos sea, en cambio, mínima, no debilita esta conclusión: la reafirma, en la medida en que se confirma que el valor predictivo reside en los determinantes identificados y en la calidad de su construcción, y no en una ventaja específica de la técnica de clasificación empleada. Con ello, los modelos construidos están demostrada y sustancialmente por encima del azar, lo que valida su empleo como insumo confiable para la identificación de perfiles de riesgo que desarrolla el resto del capítulo.

Anexo V:

Diagnóstico de Colinealidad en la Ecuación de Mincer Extendida

Este anexo documenta el análisis realizado para confirmar o descartar la existencia de colinealidad severa entre la escolaridad y el quintil de ingreso del hogar en la ecuación de Mincer extendida (subsección 4.2.2), cuya competencia por explicar una porción común de la variación del ingreso se señala como la causa de la pérdida de significancia estadística de la escolaridad en esa especificación. Se presenta, en primer lugar, el análisis del Factor de Inflación de la Varianza (VIF) sobre el conjunto completo de covariables del modelo, y en segundo lugar, la justificación de por qué no se recurrió a técnicas de regularización (Lasso o Ridge) como estrategia alternativa para tratar esta situación.

Factor de Inflación de la Varianza (VIF)

El VIF cuantifica en qué medida la varianza del coeficiente estimado de una variable se encuentra inflada como consecuencia de su correlación con el resto de covariables del modelo. Se calcula, para cada variable explicativa como:

$$VIF = 1 / (1 - R^2_j)$$

Donde: R^2_j es el coeficiente de determinación de una regresión auxiliar de esa variable sobre todas las demás covariables del modelo.

Valores de VIF cercanos a 1 indican ausencia de correlación relevante con el resto de variables; valores entre 1 y 5 se consideran aceptables y no comprometen la estabilidad de las estimaciones; valores entre 5 y 10 sugieren una colinealidad moderada que amerita atención; y valores superiores a 10 se interpretan convencionalmente como evidencia de colinealidad severa, capaz de desestabilizar los coeficientes y sus errores estándar.

El análisis se ejecutó sobre la totalidad de las 47 covariables de la ecuación de Mincer extendida, utilizando la misma matriz de diseño (X_extendido) empleada para la estimación reportada en la subsección correspondiente, sobre las 111.122 observaciones que conforman la muestra de dicha ecuación.

Cuadro V.1. Factor de Inflación de la Varianza (VIF) de las covariables de la ecuación de Mincer extendida.

Variable	**VIF**	**Categoría**
Rama: Hogares privados	2009.306	Severa (>10)
Ocupación: servicio doméstico	2009.162	Severa (>10)
Constante	49.435	Severa (>10) / No aplica
Experiencia potencial (exp_sh)	14.118	Severa (>10)
Experiencia potencial² (exp_sh²)	12.457	Severa (>10)
Rama: Comercio y vehículos	2.597	Aceptable (1-5)
Ocupación: cuenta propia	2.542	Aceptable (1-5)

Escolaridad (year_sh)	2.518	Aceptable (1-5)
Quintil de ingreso del hogar 5	2.417	Aceptable (1-5)
Ocupación: jornalero	2.289	Aceptable (1-5)
Quintil de ingreso del hogar 4	2.243	Aceptable (1-5)
Quintil de ingreso del hogar 3	2.059	Aceptable (1-5)
Informalidad	1.976	Aceptable (1-5)
Estado civil: casado/a	1.920	Aceptable (1-5)
Quintil de ingreso del hogar 2	1.888	Aceptable (1-5)
Rama: Alojamiento y comida	1.825	Aceptable (1-5)
Rama: Manufacturas	1.798	Aceptable (1-5)
Rama: Transporte	1.685	Aceptable (1-5)
Área rural	1.521	Aceptable (1-5)
Estado civil: unión libre	1.503	Aceptable (1-5)
Rama: Construcción	1.494	Aceptable (1-5)
Rama: Profesionales y técnicas	1.489	Aceptable (1-5)
Rama: Otros servicios	1.368	Aceptable (1-5)
Rama: Servicios administrativos	1.335	Aceptable (1-5)
Mujer	1.335	Aceptable (1-5)
Estado civil: viudo/a	1.329	Aceptable (1-5)
Rama: Salud y sociales	1.321	Aceptable (1-5)
Rama: Enseñanza	1.308	Aceptable (1-5)
Estado civil: separado/a	1.307	Aceptable (1-5)
Estado civil: divorciado/a	1.236	Aceptable (1-5)
Tamaño: pequeña (20-35)	1.224	Aceptable (1-5)
Ocupación: patrono	1.199	Aceptable (1-5)
Tamaño: microempresa (10-19)	1.177	Aceptable (1-5)
Rama: Información y comunicación	1.110	Aceptable (1-5)

Tamaño: mediana (50-60)	1.104	Aceptable (1-5)
Rama: Artes y recreación	1.097	Aceptable (1-5)
Tamaño: pequeña (36-49)	1.064	Aceptable (1-5)
Rama: Inmobiliarias	1.062	Aceptable (1-5)
Rama: Minas y canteras	1.058	Aceptable (1-5)
Rama: Financieras y seguros	1.055	Aceptable (1-5)
Ocupación: trabajo no remunerado	1.050	Aceptable (1-5)
Rama: Administración pública	1.030	Aceptable (1-5)
Tamaño: grande	1.029	Aceptable (1-5)
Tamaño: mediana (61-75)	1.021	Aceptable (1-5)
Rama: Agua y alcantarillado	1.014	Aceptable (1-5)
Rama: Electricidad y gas	1.005	Aceptable (1-5)
Rama: Organizaciones extraterritoriales	1.002	Aceptable (1-5)

Fuente: elaboración propia a partir de ENEMDU-INEC (2025).

El resultado central de este diagnóstico es que la escolaridad (VIF = 2,518) y las cuatro categorías del quintil de ingreso del hogar (VIF entre 1,888 y 2,417) se ubican dentro del rango convencionalmente considerado aceptable (1-5), muy por debajo del umbral de 10 que señalaría un problema estadístico grave. Este hallazgo permite descartar la colinealidad severa como explicación de la pérdida de significancia estadística de la escolaridad en la ecuación extendida: sus errores estándar no se encuentran sustancialmente inflados por correlación con el resto de covariables del modelo, incluido el quintil de ingreso del hogar. La explicación correcta, en consecuencia, no es de naturaleza estadística sino conceptual: ambas variables comparten una porción sustantiva de la variación del ingreso que explican, de modo que, una vez que el quintil de ingreso del hogar, parcialmente determinado por la escolaridad de sus miembros, ya está presente en el modelo, el aporte marginal e independiente de la escolaridad se reduce, sin que ello implique inestabilidad de las estimaciones.

El diagnóstico revela, adicionalmente, dos resultados que merecen mención. En primer lugar, los términos de experiencia potencial (VIF = 14,118) y su componente cuadrático (VIF = 12,457) sí superan el umbral de colinealidad severa, tal como es esperable dado que el segundo término es, por construcción matemática, una función del primero; esta colinealidad es mecánica y necesaria para que la ecuación capture los rendimientos decrecientes de la experiencia que predice la teoría del capital humano, y está amparada en la propia teoría de Mincer. En segundo lugar, la ocupación de servicio doméstico y la rama de hogares privados presentan un VIF extremo (2.009,306 y 2.009,162, respectivamente), reflejo de que, en la práctica, casi la totalidad de los trabajadores clasificados en la primera categoría ocupacional se emplean también en la segunda rama de actividad; se trata de una redundancia estructural entre dos

categorías específicas de la muestra, ajena a la escolaridad y al quintil de ingreso, y que no afecta la interpretación de los coeficientes de interés de este trabajo.

Así mismo destaca el VIF de la constante (49,435). Sin embargo, este coeficiente se excluye del umbral antes referido de VIF >10 con equivalencia a varianza inflada como consecuencia de su correlación con el resto de covariables del modelo; debido a que en modelos con múltiples variables categóricas codificadas por referencia, la constante no representa una variable explicativa sustantiva sobre la cual sea relevante diagnosticar colinealidad, y su VIF elevado es un resultado esperado de la parametrización del modelo.

Justificación de Por qué no se emplearon técnicas de regularización (Lasso/Ridge)

Se descartó deliberadamente el uso de Lasso o Ridge como estrategia para dar tratamiento a una posible redundancia entre las variables del modelo; con base en los resultados obtenidos en el proceso anterior y principalmente por una razón que atañe directamente al objetivo de esta investigación: Ambas técnicas de regularización introducen, por diseño, un sesgo en los coeficientes estimados: en el caso de Ridge, mediante una penalización que reduce la magnitud de todos los coeficientes hacia cero; en el caso de Lasso, mediante una penalización que puede llevar algunos coeficientes exactamente a cero, eliminándolos del modelo. Ese sesgo es intencional y deseable cuando el objetivo del ejercicio es predictivo. Sin embargo, el objetivo del presente trabajo no es maximizar la capacidad predictiva ni incrementar la proporción de la variación del ingreso explicada por el modelo, sino obtener un coeficiente de informalidad estadísticamente significativo y libre de modificaciones artificiales, amparado por un R2 regionalmente aceptado en este tipo de aplicaciones de Mincer; que permita determinar con precisión la penalización salarial asociada a esa condición. Penalizar artificialmente los coeficientes del modelo, incluido potencialmente el propio coeficiente de es_informal, introduciría una distorsión contraria a ese objetivo: convertiría una estimación insesgada, con inferencia estadística estándar plenamente válida, en una estimación sesgada por diseño, cuya magnitud dependería en parte de un parámetro de penalización elegido y no únicamente de la relación real entre las variables. Por esta razón, se optó por mantener la especificación de la Ecuación de Mincer en su versión Extendida, complementada con el diagnóstico de colinealidad aquí documentado.