

**UNIVERSIDAD PERUANA DE CIENCIAS  
APLICADAS**

**FACULTAD DE ECONOMÍA**

**PROGRAMA ACADÉMICO DE ECONOMÍA Y FINANZAS**

**EFECTOS DE LA FORMALIZACIÓN  
EMPRESARIAL SOBRE LA SOSTENIBILIDAD  
DE LAS MYPES PERUANAS EN 2022**

**TRABAJO DE INVESTIGACIÓN**

Para optar el grado de bachiller en Economía y Finanzas

**AUTOR**

León Pérez, Leonardo Abel (0009-0008-5615-6561)

**ASESOR(A)**

Castro Herrera, Soraya (0000-0002-4468-9300)

Lima, noviembre de 2025

# Índice

<b>1. INTRODUCCIÓN</b>	<b>4</b>
<b>2. HECHOS ESTILIZADOS</b>	<b>5</b>
<b>3. APROXIMACIÓN METODOLÓGICA</b>	<b>9</b>
3.1 Fuente de Datos . . . . .	9
3.2 Justificación de Variables . . . . .	10
3.3 Especificación del Modelo Econométrico . . . . .	12
3.4 Análisis de Resultados . . . . .	13
<b>4. CONCLUSIONES</b>	<b>20</b>
<b>5. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>23</b>

## Índice de tablas

1.	Descripción de variables del modelo . . . . .	11
2.	Resultados de la regresión logística . . . . .	13
3.	Odds Ratios de la regresión logística . . . . .	16

## Índice de figuras

1.	Evolución de la MYPEs Formales (2014-2021) . . . . .	5
2.	Distribución de la fuerza laboral formal/informal por tamaño de empresa (2022) . . . . .	6
3.	Ventas Totales de MYPEs, Contribución al PBI y Porcentaje de Informa- lidad (2007-2020) . . . . .	7
4.	Distribución de Empresas por Región y Sector Económico (2021) . . . .	8
5.	Distribución de Empresas por Organización Jurídica . . . . .	9
6.	Efectos Marginales de la Formalización (RUC) sobre Supervivencia por Región . . . . .	17
7.	Efectos Marginales de la Formalización según Nivel de Ventas . . . . .	18
8.	Curva ROC del Modelo Logit: Capacidad Predictiva de Supervivencia Em- presarial . . . . .	19

# 1. INTRODUCCIÓN

El desarrollo económico en economías emergentes enfrenta barreras estructurales críticas, exacerbadas por la informalidad. En Perú, las micro y pequeñas empresas (MYPEs) constituyen el 95 % del tejido empresarial y generan el 73.1 % del empleo (MTPE, 2024), pero operan bajo severas restricciones: el 75 % son informales (Torres, 2025) y solo el 6.68 % (82,094 empresas) de las microempresas y el 42.8 % (31,727 empresas) de las pequeñas empresas accedieron a financiamiento formal en 2016 (Aliaga, 2017). La pandemia de COVID-19 evidenció esta vulnerabilidad, con una mayor exposición de las unidades informales a choques externos (Varona & Gonzales, 2021). Esta fragilidad persiste: en el cuarto trimestre de 2024, la tasa de mortalidad empresarial (8.28 %) superó ampliamente a la de natalidad (2.20 %) (INEI, 2025). Frente a esto, la formalización mediante el RUC emerge como una política clave para la sostenibilidad, alineada con el ODS 8.

El marco teórico se fundamenta en el modelo de selección de Jovanovic (1982), donde las empresas descubren su eficiencia operativa ( $\theta$ ) en el mercado. La formalización actúa incrementando  $\theta$  y reduciendo costos de transacción, facilitando el acceso a mercados y crédito. Evidencia empírica respalda esto: en Perú, Chacaltana (2016) halla brechas de productividad de hasta ocho veces a favor de formales y Yamada (2009) reporta una probabilidad de cierre 15 % menor. Sin embargo, la literatura sugiere heterogeneidad geográfica: Liedholm (2002) y Tonetto et al. (2024) documentan variaciones significativas en la supervivencia según la ubicación y estructura económica regional.

En Perú, estas disparidades son marcadas. La Costa concentra el 59.67 % de MYPEs y mayor conectividad, mientras Sierra y Selva enfrentan barreras logísticas que podrían modular los beneficios de la formalización. Asimismo, la pandemia consolidó la digitalización como determinante de supervivencia (Solomon et al., 2024), donde el uso intensivo de internet se asocia a mayores utilidades (León Mendoza & Valcárcel Pineda, 2022).

Esta investigación contribuye en tres dimensiones. Primero, utiliza datos censales del V Censo Nacional Económico 2022 (1.38 millones de MYPEs), permitiendo una cobertura exhaustiva. Segundo, analiza la heterogeneidad regional mediante interacciones (RUC $\times$ Región), probando si las disparidades espaciales modulan el impacto de la formalización. Tercero, incorpora controles por regímenes tributarios y un Digital Score (0-3)

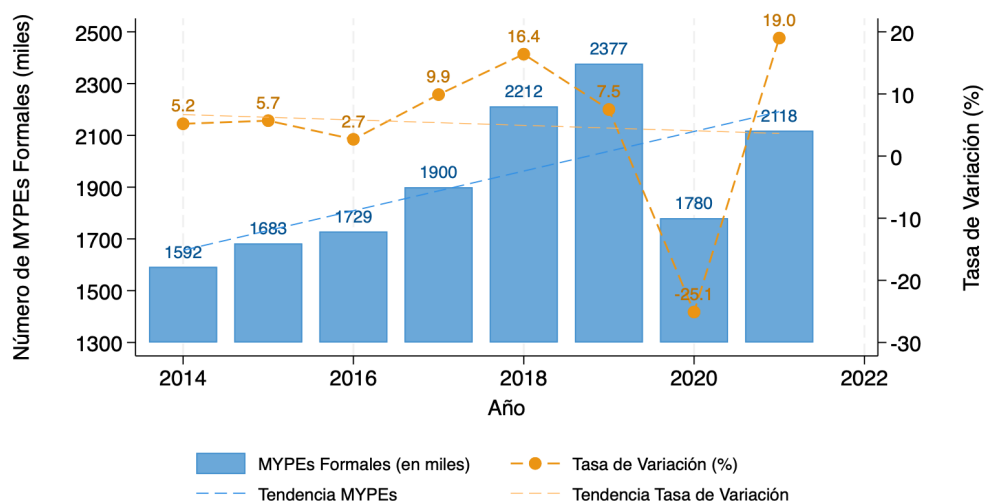
en el contexto de recuperación post-COVID-19.

El estudio aborda la pregunta: ¿Cómo incide la formalización en la supervivencia de las MYPEs peruanas y cómo varía este impacto regionalmente? Se plantea la hipótesis de que la formalización incrementa la probabilidad de supervivencia (Jovanovic, 1982), con un efecto intensificado en la Costa debido a la mayor densidad empresarial y acceso a mercados.

## 2. HECHOS ESTILIZADOS

Esta sección presenta un análisis descriptivo de las características estructurales y la dinámica reciente de las micro y pequeñas empresas (MYPEs) en el Perú, utilizando datos del V Censo Nacional Económico (INEI, 2022) y fuentes complementarias. El objetivo es contextualizar el entorno en el que operan estas firmas, identificando patrones de formalización, desempeño económico y heterogeneidad regional que fundamentan la investigación.

Figura 1: Evolución de la MYPEs Formales (2014-2021)

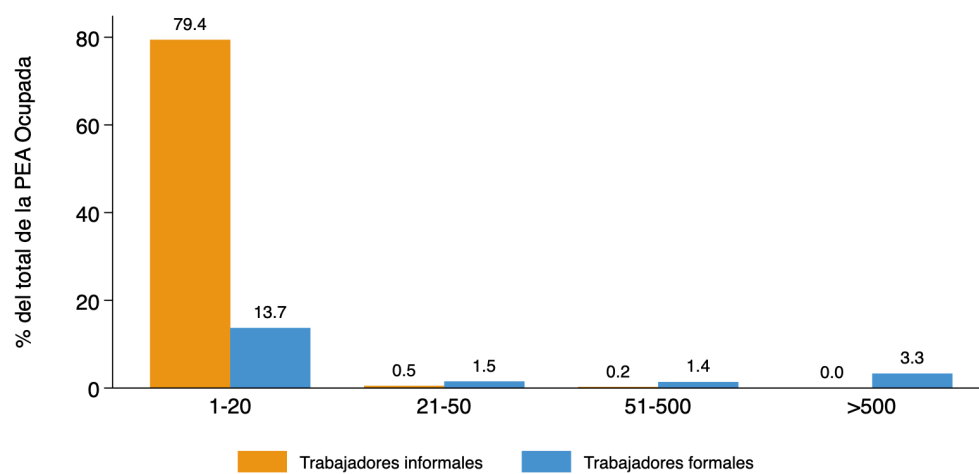


*Fuente: Elaboración propia a partir de datos del informe Las MIPYME en cifras 2017 (PRODUCE, 2023) y de la Planilla Electrónica de SUNAT (2024).*

La trayectoria de las MYPEs formales entre 2014 y 2021 evidencia una expansión acumulada del 33.1 %, alcanzando un total de 2.12 millones de unidades registradas (INEI, 2022). No obstante, este crecimiento nominal enmascara una desaceleración estructural

en la tasa de formalización anual, tendencia que se vio drásticamente exacerbada por la crisis sanitaria de 2020, año en que se registró una contracción neta del 25.1 % en el parque empresarial (Varona & Gonzales, 2021). Este comportamiento cíclico es consistente con la dinámica de selección de mercado descrita por Jovanovic (1982), donde las firmas menos eficientes optan por la salida o la desformalización como estrategia de supervivencia ante costos regulatorios y choques exógenos adversos. La persistencia de altas tasas de salida del sector formal sugiere que las barreras burocráticas y la baja productividad continúan limitando la sostenibilidad de la formalización.

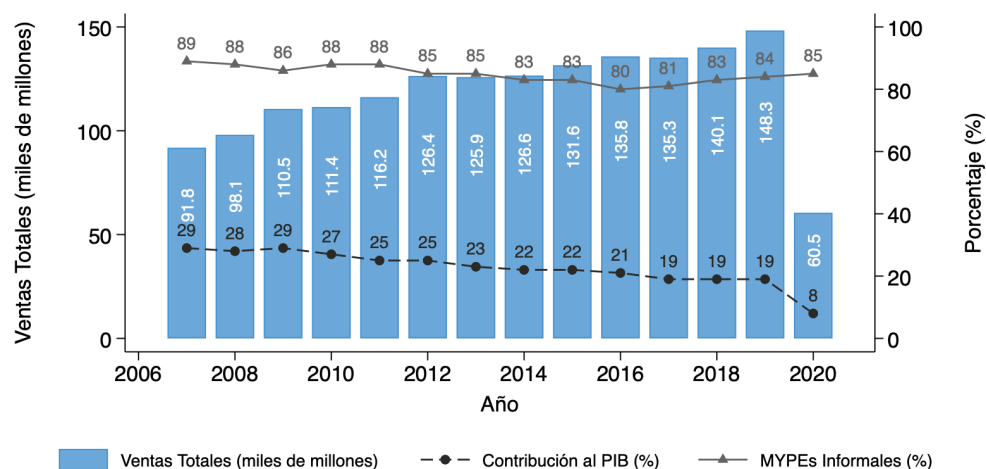
Figura 2: Distribución de la fuerza laboral formal/informal por tamaño de empresa (2022)



*Fuente: Elaboración propia a partir de datos de la ENAHO (INEI, 2020) y MTPE (2023).*

La estructura del mercado laboral peruano se caracteriza por una alta concentración en unidades económicas de baja escala: las micro y pequeñas empresas absorben conjuntamente el 95.1 % de la fuerza laboral total, abarcando tanto el sector formal como el informal (MTPE, 2024). Esta atomización del empleo tiene implicancias directas sobre la productividad agregada, dado que la gran mayoría de estos trabajadores se insertan en unidades con limitadas economías de escala. Como señala Chacaltana (2016), la predominancia de la microempresa está correlacionada con menores niveles de inversión en capital humano y tecnológico, lo que perpetúa círculos viciosos de baja productividad y alta vulnerabilidad laboral, restringiendo las posibilidades de movilidad hacia segmentos de mayor valor agregado.

Figura 3: Ventas Totales de MYPEs, Contribución al PBI y Porcentaje de Informalidad (2007-2020)

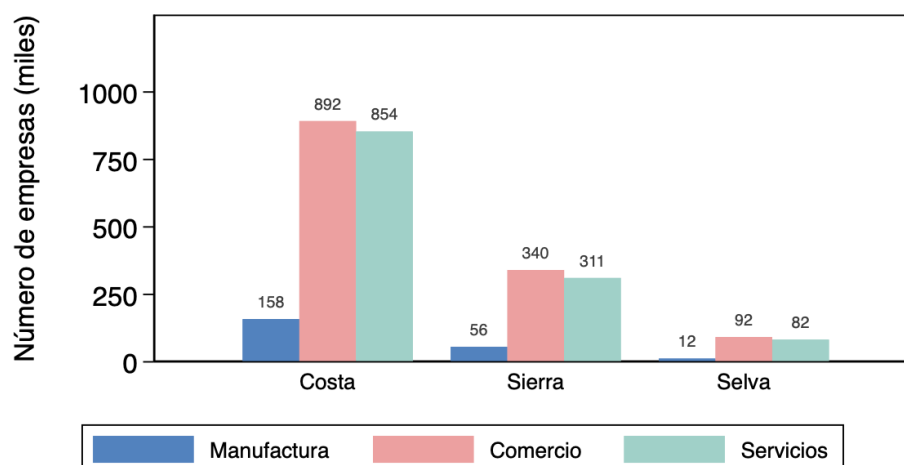


Fuente: Elaboración propia a partir de datos de la ENAHO (INEI, 2020).

A pesar de que el volumen de ventas agregado de las MYPEs experimentó un incremento sustancial del 61.5 % entre 2007 y 2019, su participación relativa en el Producto Bruto Interno ha mostrado una tendencia decreciente, cayendo drásticamente al 8 % durante la crisis de 2020 (INEI, 2022). Esta desconexión entre el dinamismo comercial y la generación de valor agregado se agrava por la persistencia de la informalidad, que repuntó al 85 % en el contexto de la pandemia. Este panorama evidencia la fragilidad estructural del sector para sostener niveles de competitividad en el largo plazo, sugiriendo que el crecimiento de las ventas no se ha traducido proporcionalmente en mejoras de eficiencia o formalización sostenida.



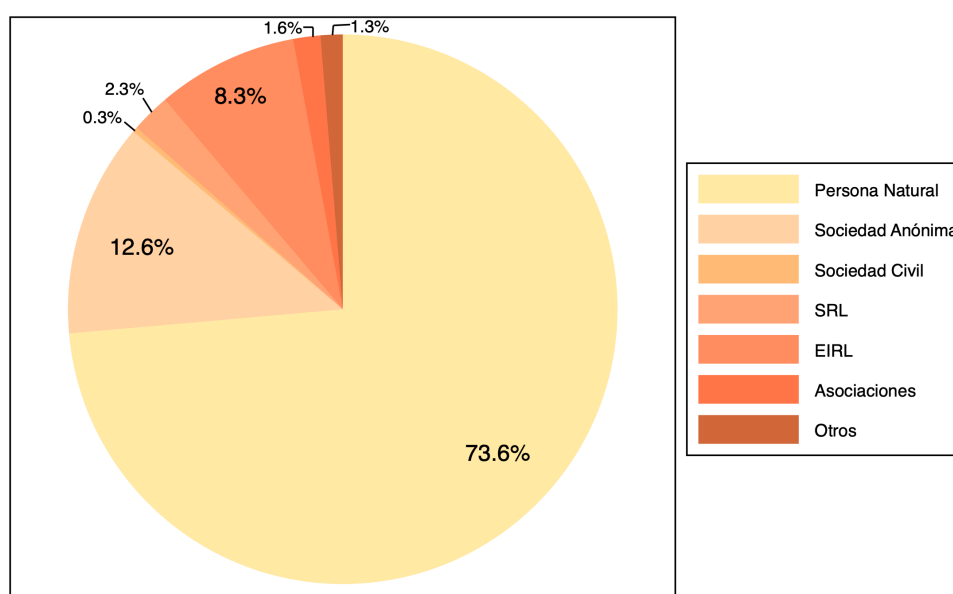
Figura 4: Distribución de Empresas por Región y Sector Económico (2021)



*Fuente: Elaboración propia a partir de datos de Perú: Estructura empresarial, 2021 (INEI, 2024).*

El análisis sectorial revela una predominancia transversal del comercio y los servicios en las tres macrorregiones del país, relegando a la actividad manufacturera a un rol marginal (INEI, 2022). Sin embargo, esta aparente uniformidad sectorial coexiste con profundas disparidades en infraestructura y acceso a mercados entre la Costa, Sierra y Selva. Como sugiere Liedholm (2002), estas asimetrías regionales en el entorno de negocios constituyen determinantes críticos de la supervivencia empresarial, otorgando ventajas comparativas a las firmas ubicadas en zonas urbanas costeras —donde se concentran las economías de aglomeración— frente a sus pares en regiones con menor densidad económica y mayores costos logísticos.

Figura 5: Distribución de Empresas por Organización Jurídica



*Fuente: Elaboración propia a partir de datos de Perú: Estructura empresarial, 2021 (INEI, 2024).*

Desde una perspectiva jurídica, el tejido empresarial peruano presenta una marcada preferencia por la figura de Persona Natural, categoría que agrupa al 73.6 % de las empresas registradas en 2021 (INEI, 2022). Esta configuración legal está intrínsecamente vinculada a la microempresa y refleja una estructura de propiedad predominantemente unipersonal. Si bien esta modalidad facilita la constitución inicial y reduce costos administrativos, puede imponer limitaciones significativas para la expansión corporativa, la separación de patrimonios y el acceso a instrumentos financieros sofisticados, condicionando las perspectivas de crecimiento y supervivencia de la firma en el mediano plazo.

### 3. APROXIMACIÓN METODOLÓGICA

#### 3.1 Fuente de Datos

El análisis empírico utiliza datos transversales del V Censo Nacional Económico 2022, (Instituto Nacional de Estadística e Informática [INEI], 2022), correspondientes al ejercicio fiscal 2021. Tras filtrar la base original para excluir establecimientos no particulares, medianas y grandes empresas, y observaciones con datos inconsistentes, la muestra final contiene 1,377,931 MYPEs: 96.6 % microempresas y 3.4 % pequeñas empresas, distribui-

das en 59.67 % Costa, 32.47 % Sierra y 7.87 % Selva. La productividad laboral se calcula como valor agregado por trabajador ocupado, siguiendo la metodología estándar del INEI (2022). El análisis se implementa en Stata 17 (StataCorp, 2021).

### 3.2 Justificación de Variables

La variable dependiente es supervivencia empresarial (0=no operativa, 1=operativa en 2021). La variable independiente principal es formalización mediante tenencia de RUC (0=sin RUC, 1=con RUC), representando el mecanismo teórico central que reduce costos de transacción  $c(\cdot)$  e incrementa eficiencia operativa  $\theta$  según Jovanovic (1982).

Las variables de control capturan tres dimensiones críticas. Primero, características geográficas: región (0=Costa, 1=Sierra, 2=Selva) modula efectos de formalización mediante acceso diferencial a mercados e infraestructura institucional. Los términos de interacción  $RUC \times Región$  testean la hipótesis de heterogeneidad espacial documentada, entre otros autores, por Liedholm (2002).

Segundo, características económicas y productivas. Digital Score (0-3) captura intensidad de adopción digital post-COVID mediante presencia web, Facebook y redes sociales, variable crítica dado el contexto temporal de 2021 donde digitalización se transformó en factor de supervivencia. Productividad laboral (valor agregado por trabajador) representa la eficiencia operativa  $\theta$  del modelo de Jovanovic, distinguiendo empresas viables de no viables. Ventas anuales refleja capacidad de generar beneficios  $\pi_t$  y escala operativa; se escaló dividiéndola por 1,000 (expresada en miles de soles) para resolver problemas de convergencia numérica en algoritmos de optimización sin alterar propiedades estadísticas del modelo.

Régimen tributario (RUS, RER, RG, RMT) captura intensidad regulatoria dentro de empresas formales, controlando efectos heterogéneos de carga administrativa según sofisticación del régimen. Variables demográficas y operacionales adicionales (sector, sexo del gerente, tributos, remuneraciones, tipo de local) se incluyeron como controles siguiendo literatura previa, aunque resultaron no significativas estadísticamente en el modelo final.

Tamaño empresarial se excluye del modelo porque en Perú se define por ventas anuales según Ley 30056 (Congreso de la República del Perú, 2013), generando colinealidad con

la variable ventas incluida. La muestra presenta 96.6 % microempresas y 3.4 % pequeñas empresas, desbalance controlado mediante errores estándar clusterizados por código CIU de 4 dígitos. Variables de tasa de interés se descartan por ausencia de datos censales sobre acceso individual a préstamos y problemas de datos faltantes no aleatorios (solo 6.68 % de MYPEs accede a financiamiento formal según Aliaga, 2017). La Tabla 1 resume todas las variables utilizadas.

Tabla 1: Descripción de variables del modelo

Variable	Definición	Codificación
Supervivencia	Operativa en 2021	Binaria: 0=No, 1=Sí
RUC	Tenencia de Registro Único de Contribuyentes	Binaria: 0=Sin RUC, 1=Con RUC
Región	Ubicación geográfica	0=Costa, 1=Sierra, 2=Selva
Sector	Sector económico	0=Comercial, 1=Servicios, 2=Productivo
Sexo Gerente	Género del gerente	Binaria: 0=Mujer, 1=Hom- bre
Ventas Netas	Ingresos netos 2021 (miles de so- les)	Continua
Productividad	Valor agregado por trabajador (soles)	Continua
Digital Score	Intensidad digitalización	Ordinal: 0-3 instrumentos digitales
Tributos	Tributos pagados 2021 (soles)	Continua
Remuneraciones	Salarios y beneficios 2021 (soles)	Continua
Tipo Local	Tenencia del local	0=Propio, 1=Alquilado, 2=Otro
Régimen	Régimen Tributario	0=RUS, 1=RER, 2=RG, 3=RMT

*Fuente: V Censo Nacional Económico 2022 (INEI), elaboración propia.*

### 3.3 Especificación del Modelo Econométrico

El análisis empírico emplea un modelo de regresión logística para estimar la probabilidad de supervivencia de las MYPEs, lo cual se implementó en Stata 17 (StataCorp, 2021), software estándar en econometría aplicada. La ecuación logística es:

$$\begin{aligned} \ln \left( \frac{P(Y = 1)}{1 - P(Y = 1)} \right) = & \beta_0 + \beta_1 RUC + \beta_2 Sierra + \beta_3 Selva \\ & + \beta_4 (RUC \times Sierra) + \beta_5 (RUC \times Selva) \\ & + \beta_6 Sector + \beta_7 Genero + \beta_8 Ventas_{2021} \\ & + \beta_9 Productividad + \beta_{10} DigitalScore \\ & + \beta_{11} Tributos + \beta_{12} Remuneraciones \\ & + \beta_{13} TipoLocal + \beta_{14} Regimen + \varepsilon \end{aligned}$$

donde  $Y=1$  indica supervivencia (operativa en 2021), RUC es formalización (binaria), Sierra y Selva son variables dummy regionales (Costa como base), los términos de interacción  $RUC \times Sierra$  y  $RUC \times Selva$  capturan efectos heterogéneos regionales de la formalización, y las demás variables son controles empresariales y demográficos. Los errores estándar se agrupan por código CIIU de 4 dígitos para controlar heterogeneidad sectorial no observada, siguiendo la recomendación de Cameron & Miller (2015).

La interpretación de coeficientes de interacción requiere atención especial porque las interacciones frecuentemente generan confusión. El coeficiente  $\beta_1$  captura el efecto de RUC en Costa, la región base. El coeficiente  $\beta_4$  representa la diferencia en el efecto de RUC entre Sierra y Costa, mientras que  $\beta_5$  representa la diferencia entre Selva y Costa. El efecto total de RUC en cada región se calcula como: Costa =  $\beta_1$ ; Sierra =  $\beta_1 + \beta_4$ ; Selva =  $\beta_1 + \beta_5$ . Para testear si el efecto de RUC difiere significativamente entre regiones, se emplean comandos de prueba de hipótesis en Stata, permitiendo determinar si la heterogeneidad regional es estadísticamente significativa.

La elección del modelo logit sobre alternativas como probit o cloglog se justifica por tres razones. Primero, la distribución logística facilita interpretación de coeficientes como logaritmos de odds ratios, con transformación directa a odds ratios mediante la función exponencial. Segundo, el modelo permite calcular efectos marginales mediante el coman-

do margins en Stata, facilitando interpretación en unidades de probabilidad que son más intuitivas para recomendaciones de política. Tercero, la especificación con términos de interacción permite testear formalmente la hipótesis de heterogeneidad regional, superando regresiones separadas por región que no permiten comparación estadística directa y reducen eficiencia al no aprovechar información conjunta de toda la muestra. La hipótesis central se valida si  $\beta$  es positivo y significativo ( $p < 0.05$ ), indicando que formalización incrementa supervivencia en Costa, y si  $\beta_1$  y  $\beta_2$  son negativos y significativos, confirmando efectos menores en Sierra y Selva como predice el modelo teórico dado el mejor desarrollo institucional costero.

### 3.4 Análisis de Resultados

El modelo logit estimado incorpora 1,327,956 observaciones con ajuste global estadísticamente significativo (Wald  $\chi^2=1581.95$ ,  $p < 0.0001$ ). El Pseudo- $R^2$  de McFadden de 0.0463 es consistente con literatura metodológica sobre modelos logísticos aplicados a fenómenos con alta heterogeneidad no observable (Hemmert et al., 2016). La prueba VIF confirma ausencia de multicolinealidad problemática (VIF promedio=1.20, máximo=1.57 para ventas), validando la inclusión simultánea de controles correlacionados. Errores estándar clusterizados por código CIIU de 4 dígitos controlan correlación intra-industrial.

La Tabla 2 presenta resultados del modelo logit, revelando efectos contraintuitivos de formalización sobre supervivencia. El coeficiente RUC en Costa ( $\beta = -0.2161$ ,  $SE=0.0802$ ,  $z=-2.70$ ,  $p=0.007$ ) es negativo y estadísticamente significativo, correspondiendo a un odds ratio de 0.8057 (IC 95 %: 0.688-0.943). Esto indica que empresas formales en Costa reducen sus chances de supervivencia en 19.4 % versus informales, manteniendo constantes las demás variables. Este resultado contradice la hipótesis inicial pero revela dinámicas complejas de formalización en contexto post-COVID.

Tabla 2: Resultados de la regresión logística

Variable dependiente: op2021_original					
Variable	Coef.	Robust SE	$z$	$p$ -value	95 % CI
<b>RUC</b>	-0.216	0.080	-2.70	0.007	[-0.373, -0.059]

*Continúa en la siguiente página*

Tabla 2 (continuación)

Variable	Coef.	Robust SE	$z$	$p$ -value	95 % CI
<i>Región (Base: Costa)</i>					
Sierra	-0.423***	0.059	-7.12	0.000	[-0.540, -0.307]
Selva	-0.134***	0.040	-3.39	0.001	[-0.212, -0.057]
<i>Región <math>\times</math> RUC Interacción</i>					
Sierra $\times$ RUC	0.262***	0.074	3.53	0.000	[0.116, 0.408]
Selva $\times$ RUC	-0.013	0.063	-0.20	0.841	[-0.136, 0.111]
<b>Ventas (miles)</b>	.0003***	0.000	6.54	0.000	[0.000, 0.000]
<b>Productividad (miles)</b>	0.013***	0.002	6.58	0.000	[0.009, 0.017]
<b>Digital Score</b>	-0.111***	0.024	-4.72	0.000	[-0.157, -0.065]
<i>Régimen tributario</i>					
Nuevo RUS	0.283***	0.040	7.10	0.000	[0.205, 0.361]
RER	0.230***	0.064	3.58	0.000	[0.104, 0.356]
RMT	0.612***	0.079	7.79	0.000	[0.458, 0.766]
Régimen General	1.158***	0.179	6.46	0.000	[0.806, 1.509]
<b>Constant</b>	0.146	0.113	1.29	0.196	[-0.075, 0.367]
Observations	1,327,956				
Clusters (ciiu_4dig)	335				
Wald $\chi^2(19)$	1581.95				
Prob $> \chi^2$	0.0000				
Pseudo $R^2$	0.0463				
Log pseudolikelihood	-877,190.14				

*Notas:* Errores estándares robustos (clusterizados por CIIU de 4 dígitos).

Los términos de interacción revelan heterogeneidad regional crítica. RUC $\times$ Sierra ( =0.2622, SE=0.0744, z=3.53, p<0.001) es positivo y altamente significativo, indicando que el efecto de formalización en Sierra difiere sustancialmente de Costa. El efecto total en Sierra es  $\beta_1 + \beta_2 =$ -

$0.2161+0.2622=0.0461$  (OR=1.0472), prácticamente neutral o levemente positivo. RUC×Selva ( $=-0.0126$ ,  $p=0.841$ ) no alcanza significancia, confirmando que el efecto en Selva es estadísticamente indistinguible de Costa. Tests de heterogeneidad confirman diferencias regionales: prueba conjunta de interacciones rechaza efectos homogéneos ( $\chi^2=12.81$ ,  $p=0.0016$ ); comparación Sierra versus Costa es altamente significativa ( $\chi^2=12.43$ ,  $p=0.0004$ ); comparación Selva versus Costa no significativa ( $\chi^2=0.04$ ,  $p=0.841$ ).

Las variables de control exhiben signos teóricamente esperados, validando la especificación del modelo. Productividad laboral (OR=1.0132 por mil soles,  $p<0.001$ ) y ventas (OR=1.0003,  $p<0.001$ ) incrementan supervivencia, consistente con el modelo de Jovanovic donde mayor eficiencia y escala operativa mejoran rentabilidad. Digital Score muestra efecto negativo contraintuitivo (OR=0.8950 por unidad,  $p<0.001$ ), posiblemente reflejando que empresas en dificultades adoptaron estrategias digitales reactivas durante la pandemia sin lograr mejoras inmediatas de rentabilidad. Régimen tributario presenta efectos fuertemente positivos: RUS incrementa chances en 32.7%, RER en 25.9%, RMT en 84.5%, Régimen General en 218.3%, sugiriendo que empresas con capacidad de operar bajo regímenes más sofisticados poseen características organizacionales superiores. Variables como sector, sexo del gerente, tributos y remuneraciones resultaron no significativas estadísticamente.

La Tabla 3 presenta efectos marginales de RUC por región, calculados mediante el comando `margins` en Stata para interpretación económica directa. En Costa, poseer RUC reduce la probabilidad de supervivencia en -5.09 puntos porcentuales (SE=1.88,  $p=0.007$ ), rechazando la hipótesis de efectos positivos de formalización en esta región. En Sierra, el efecto marginal es positivo pero no significativo (+1.08 pp, SE=0.89,  $p=0.224$ ), sugiriendo efecto neutral. En Selva, el efecto marginal es -5.41 puntos porcentuales (SE=2.24,  $p=0.016$ ), similar a Costa y consistente con la no significancia de la interacción Selva×RUC. Tests de diferencias confirman que el efecto en Sierra difiere significativamente de Costa ( $\chi^2=12.43$ ,  $p=0.0004$ ) y Selva ( $\chi^2=9.01$ ,  $p=0.0027$ ), mientras que Costa y Selva no difieren ( $\chi^2=0.04$ ,  $p=0.841$ ). Estos hallazgos rechazan ambas hipótesis específicas: ni la formalización incrementa uniformemente supervivencia, ni los efectos son más pronunciados en Costa.



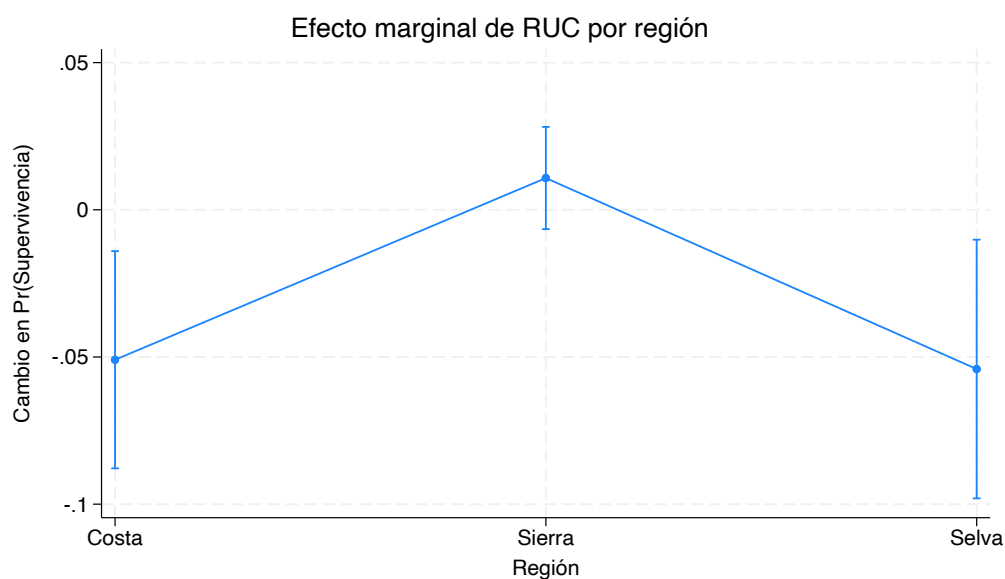
Tabla 3: Odds Ratios de la regresión logística

Variable	Variable dependiente: op2021_original				
	Odds Ratio	Robust SE	$z$	$p$ -value	95 % CI
<b>RUC</b>	0.806***	0.065	−2.70	0.007	[0.689, 0.943]
<i>Región (Base: Costa)</i>					
Sierra	0.655***	0.039	−7.12	0.000	[0.583, 0.736]
Selva	0.874***	0.035	−3.39	0.001	[0.809, 0.945]
<i>Región × RUC Interacción</i>					
Sierra × RUC	1.300***	0.097	3.53	0.000	[1.124, 1.504]
Selva × RUC	0.987	0.062	−0.20	0.841	[0.873, 1.117]
<b>Ventas (miles)</b>	1.0003***	0.000	6.54	0.000	[1.000, 1.000]
<b>Productividad (miles)</b>	1.013***	0.002	6.58	0.000	[1.009, 1.017]
<b>Digital Score</b>	0.895***	0.021	−4.72	0.000	[0.855, 0.937]
<i>Régimen tributario</i>					
Nuevo RUS	1.327***	0.053	7.10	0.000	[1.227, 1.435]
RER	1.259***	0.081	3.58	0.000	[1.110, 1.428]
RMT	1.845***	0.145	7.79	0.000	[1.581, 2.152]
Régimen General	3.183***	0.571	6.46	0.000	[2.240, 4.524]
<b>Constante</b>	1.157	0.130	1.29	0.196	[0.928, 1.443]
Observations	1,327,956				
Clusters (ciiu_4dig)	335				
Wald $\chi^2(19)$	1581.95				
Prob > $\chi^2$	0.0000				
Pseudo $R^2$	0.0463				
Log pseudolikelihood	−877, 190.14				

*Notas:* Errores estándar robustos agrupados a nivel CIIU 4 dígitos (335 clusters).

El patrón de resultados contradice predicciones teóricas convencionales, pero revela dinámicas complejas del proceso de formalización en contextos de alta informalidad estructural. Tres explicaciones complementarias reconcilian estos hallazgos con el marco teórico. Primero, el contexto temporal de 2021 como año de recuperación post-COVID introduce distorsiones donde costos inmediatos de cumplimiento superan beneficios de mediano plazo, especialmente cuando solo 6.68 % de MYPEs acceden a financiamiento formal (Aliaga, 2017). Segundo, selección adversa no observable puede generar causalidad reversa si empresas con menor viabilidad intrínseca formalizan reactivamente para acceder a programas de emergencia (Reactiva Perú, FAE-MYPE). Tercero, el horizonte temporal captura únicamente supervivencia de corto plazo, insuficiente para que beneficios estratégicos de formalización (reputación, contratos públicos, expansión interregional) se materialicen completamente. La Figura 1 ilustra gráficamente estos efectos marginales por región, evidenciando la heterogeneidad espacial documentada.

Figura 6: Efectos Marginales de la Formalización (RUC) sobre Supervivencia por Región



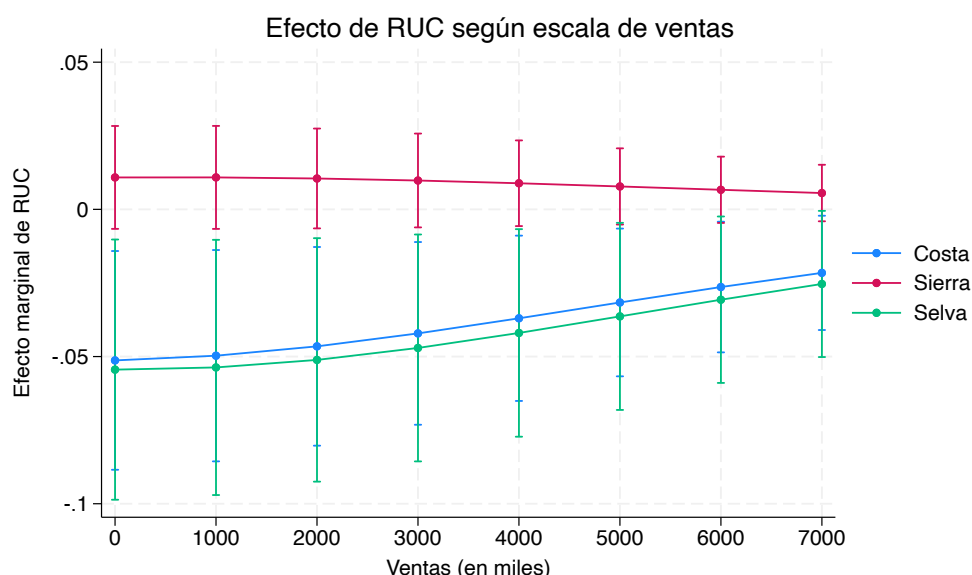
*Fuente: Elaboración propia con datos del V Censo Nacional Económico 2022 (INEI). Nota: Las barras representan el incremento en puntos porcentuales de la probabilidad de supervivencia asociado a la posesión de RUC, evaluado en valores medios de las variables de control. Intervalos de confianza al 95 %.*

La heterogeneidad regional refleja diferencias estructurales en mercados laborales, densidad empresarial e infraestructura institucional. En Costa, donde se concentra 66.5 % de MYPEs formales, la competencia intensa en mercados saturados magnifica costos relativos de cumplimiento regulatorio sin garantizar ventajas competitivas diferenciadas, explicando el efecto negativo. En

Sierra, donde solo 26.2 % de MYPEs operan, la formalización podría funcionar como señal de seriedad empresarial en mercados menos competitivos, atenuando penalidades observadas en otras regiones, aunque limitaciones de infraestructura digital (39 % acceso internet en áreas urbanas serranas versus 59 % en Lima, Instituto Nacional de Estadística e Informática [INEI], 2020) impiden que este efecto señalizador alcance significancia robusta. En Selva, barreras logísticas y dispersión poblacional neutralizan tanto costos como beneficios, replicando patrones negativos similares a Costa.

El análisis de efectos marginales según niveles de ventas revela no linealidades importantes. La Figura 2 muestra cómo el efecto marginal de la formalización varía según la escala operativa de la empresa, sugiriendo que los costos del RUC son heterogéneos no solo geográficamente sino también según tamaño empresarial.

Figura 7: Efectos Marginales de la Formalización según Nivel de Ventas



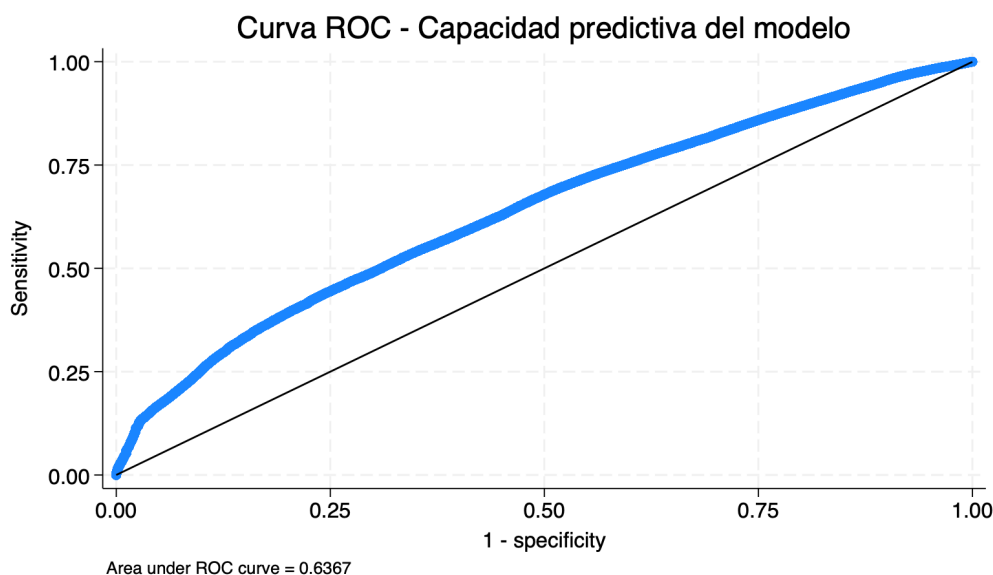
*Fuente: Elaboración propia con datos del V Censo Nacional Económico 2022 (INEI). Nota: El gráfico muestra la relación no lineal entre ventas anuales y el efecto marginal de RUC sobre supervivencia, calculado mediante el comando margins en Stata. La línea representa valores predichos y el área sombreada intervalos de confianza al 95 %.*

Respecto al régimen tributario, los coeficientes positivos y significativos observados (RUS: OR=1.327, RER: OR=1.259, RMT: OR=1.845, Régimen General: OR=3.183) requieren interpretación cautelosa debido a potenciales problemas de simultaneidad. Es plausible que empresas con mayor viabilidad intrínseca, derivada de características no observables como capacidad gerencial, acceso a redes empresariales o eficiencia organizacional, autoseleccionen

regímenes más complejos precisamente porque poseen la capacidad operativa para absorber los costos administrativos asociados. En este escenario, el efecto positivo capturado no refleja necesariamente que el régimen tributario per se incremente la supervivencia, sino que empresas inherentemente más productivas optan por regímenes sofisticados. La dirección causal requiere métodos de identificación adicionales, como variables instrumentales o diseños cuasi-experimentales que permitan aislar el efecto causal del régimen sobre supervivencia del efecto de autoselección.

Finalmente, la capacidad predictiva del modelo se evalúa mediante curvas ROC (Receiver Operating Characteristic). La Figura 3 presenta la curva ROC del modelo logit con interacciones regionales, evidenciando capacidad discriminatoria moderada con área bajo la curva de 0.6367. Este resultado indica que el modelo captura parcialmente los factores determinantes de supervivencia empresarial, sugiriendo que variables no observables desempeñan un rol relevante en el contexto peruano.

Figura 8: Curva ROC del Modelo Logit: Capacidad Predictiva de Supervivencia Empresarial



*Fuente: Elaboración propia con datos del V Censo Nacional Económico 2022 (INEI). Nota: La curva ROC evalúa el trade-off entre sensibilidad (tasa de verdaderos positivos) y especificidad (1 - tasa de falsos positivos) para diferentes umbrales de clasificación. El área bajo la curva (AUC) de 0.6367 indica capacidad discriminatoria moderada del modelo. La línea diagonal representa desempeño de un clasificador aleatorio (AUC=0.50).*

## 4. CONCLUSIONES

El análisis econométrico con 1,327,956 MYPEs revela hallazgos que contradicen parcialmente las hipótesis iniciales. Contrario a predicciones teóricas del modelo de Jovanovic (1982), la formalización mediante RUC reduce la probabilidad de supervivencia en Costa (-5.09 pp,  $p < 0.05$ ) y Selva (-5.41 pp,  $p < 0.05$ ). En Sierra se observa efecto positivo no significativo (+1.08 pp,  $p = 0.224$ ), aunque la prueba de heterogeneidad confirma diferencias estadísticamente significativas respecto a Costa ( $\chi^2 = 12.43$ ,  $p = 0.0004$ ). La prueba conjunta de interacciones regionales rechaza efectos homogéneos ( $\chi^2 = 12.81$ ,  $p = 0.0016$ ), validando heterogeneidad territorial sustancial. Estos resultados rechazan ambas hipótesis específicas: ni la formalización incrementa uniformemente supervivencia, ni los efectos son más pronunciados en Costa.

Tres mecanismos complementarios permiten reconciliar la aparente contradicción entre predicciones teóricas y hallazgos empíricos. En primer lugar, el año 2021 representa un período de recuperación post-pandemia en el cual las cargas administrativas y fiscales inmediatas de la formalización exceden los retornos esperables en el mediano plazo, especialmente considerando que únicamente 6.68 % de las MYPEs logra acceder a crédito formal (Aliaga, 2017), lo cual restringe la concreción de beneficios potenciales asociados al estatus formal. En segundo lugar, dinámicas de selección adversa no observables pueden inducir causalidad inversa: si las firmas con menores perspectivas de viabilidad optan por formalizarse como estrategia reactiva para calificar a programas de asistencia estatal (Reactiva Perú, FAE-MYPE), el efecto causal del RUC se confunde con atributos preexistentes de empresas vulnerables. En tercer lugar, la ventana temporal analizada (ejercicio fiscal 2021) captura exclusivamente efectos de corto plazo, período insuficiente para que ventajas estratégicas de la formalización (construcción reputacional, participación en contrataciones públicas, expansión geográfica) logren materializarse plenamente. Como señala Jovanovic (1982), el parámetro de eficiencia  $\theta$  se descubre progresivamente mediante la interacción empresarial con mercados, dinámica incompatible con el análisis de corte transversal limitado a un año posterior a la crisis.

Los patrones de heterogeneidad regional observados responden a disparidades estructurales en configuraciones de mercados laborales, concentración empresarial y desarrollo institucional. La región Costa, que alberga 66.5 % de las MYPEs formalizadas, exhibe mercados altamente competitivos y saturados donde las obligaciones regulatorias incrementan los costos operativos relativos sin generar diferenciación competitiva sustancial, fenómeno que explica el efecto negativo identificado. En la Sierra, donde apenas 26.2 % de las MYPEs opera, la formalización podría operar como mecanismo de señalización de solvencia empresarial en entornos menos competi-

vos con baja densidad de firmas registradas, mitigando así las penalidades detectadas en otras macrorregiones; sin embargo, las carencias en infraestructura digital (39 % de penetración de internet en zonas urbanas serranas frente a 59 % en Lima, según el Instituto Nacional de Estadística e Informática [INEI], 2020) obstaculizan que este efecto señalizador alcance robustez estadística. En la Selva, las severas restricciones logísticas y la elevada dispersión demográfica anulan simultáneamente tanto las desventajas como las potenciales ventajas de la formalización, reproduciendo efectos negativos análogos a los de la Costa.

Las variables de control exhiben signos teóricamente esperados, validando la especificación del modelo. Productividad laboral ( $OR=1.0132$ ,  $p<0.001$ ) y ventas ( $OR=1.0003$ ,  $p<0.001$ ) incrementan supervivencia consistentemente con el modelo de selección de Jovanovic, donde mayor eficiencia  $\theta$  y escala operativa mejoran rentabilidad  $\pi$ . Digital Score muestra efecto negativo contraintuitivo ( $OR=0.8950$ ,  $p<0.001$ ), posiblemente reflejando que empresas en dificultades adoptaron estrategias digitales reactivas durante la pandemia sin lograr mejoras inmediatas de rentabilidad. Régimen tributario presenta efectos fuertemente positivos, aunque estos coeficientes requieren interpretación cautelosa debido a problemas potenciales de simultaneidad: empresas con mayor viabilidad intrínseca pueden autoseleccionarse en regímenes más complejos porque poseen capacidad operativa para absorber costos administrativos asociados. Variables como sector, sexo del gerente, tributos y remuneraciones resultaron no significativas estadísticamente. El ajuste global es estadísticamente significativo (Wald  $\chi^2=1581.95$ ,  $p<0.0001$ ) con área bajo curva ROC de 0.6367, indicando capacidad discriminatoria moderada del modelo. El pseudo  $R^2$  de 0.0463, aunque modesto, es consistente con literatura metodológica que documenta valores sistemáticamente bajos en modelos logísticos aplicados a fenómenos con alta heterogeneidad no observable (Hemmert et al., 2018), especialmente en muestras grandes donde significancia estadística no implica pseudo  $R^2$  elevados.

La principal restricción metodológica deriva de la naturaleza transversal del V Censo Nacional Económico 2022, que imposibilita inferencias causales definitivas sobre la relación entre formalización y supervivencia. La captura del estado operativo durante un único ejercicio fiscal (2021) excluye la posibilidad de documentar trayectorias longitudinales de las firmas, transiciones en el estatus de formalización o acumulación temporal de beneficios. Sin embargo, el censo representa la única fuente de datos que incorpora simultáneamente unidades formales e informales con desagregación a nivel microeconómico individual, aventajando a registros administrativos que únicamente documentan empresas formalizadas o a estadísticas agregadas regionalmente que imposibilitan el análisis de heterogeneidad microeconómica exigido por el marco conceptual de Jovanovic (1982).

El horizonte temporal de supervivencia (año fiscal 2021) captura únicamente efectos de corto plazo, insuficiente para que beneficios estratégicos de formalización (reputación empresarial, acceso a contratos públicos, expansión interregional) se materialicen. Esta limitación es particularmente relevante dado el contexto post-COVID-19, donde distorsiones macroeconómicas pueden dominar efectos microeconómicos de formalización. Finalmente, el desbalance muestral hacia microempresas (96.6 %) refleja fielmente la composición del tejido empresarial peruano, pero limita generalización hacia pequeñas empresas más consolidadas. Errores estándar clustered por sector mitigan heterogeneidad no observada, pero análisis estratificado podría revelar dinámicas diferenciadas que el modelo agregado enmascara.

Los hallazgos demandan reformas estructurales en estrategias de formalización, contribuyendo al Objetivo de Desarrollo Sostenible 8. Primero, abandonar enfoques nacionales homogéneos en favor de políticas regionales diferenciadas. En Costa, donde formalización reduce supervivencia (-5.09 pp), la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) y el Ministerio de Producción deben reducir costos administrativos mediante ventanillas únicas digitales y simplificación tributaria, equilibrando la ecuación costo-beneficio. En Sierra, donde efectos son neutrales, intensificar acompañamiento técnico post-formalización y subsidiar conectividad digital para potenciar el efecto señalizador de RUC en mercados menos saturados. En Selva, inversiones en infraestructura logística son prerrequisito para que formalización genere beneficios, dado que dispersión poblacional neutraliza actualmente tanto costos como ventajas.

Segundo, transformar formalización de trámite administrativo a intervención integral. Dado que solo 6.68 % de MYPEs acceden a financiamiento formal (Aliaga, 2017), implementar líneas de crédito condicionadas a RUC con tasas preferenciales y períodos de gracia permitiría materializar beneficios inmediatos. Capacitación en gestión financiera y contabilidad formal debe acompañar obligatoriamente el registro, permitiendo capitalizar oportunidades que brinda el estatus formal. La evidencia de selección adversa exige focalizar apoyos en empresas con viabilidad de mediano plazo, implementando sistemas de evaluación que distingan formalización proactiva (estrategia de crecimiento) de formalización reactiva (respuesta a crisis). Incentivos indiscriminados atraen principalmente empresas en dificultades, explicando parcialmente los efectos negativos observados.

Finalmente, inversiones complementarias en infraestructura digital y acceso a mercados son condiciones necesarias para que formalización genere efectos positivos en regiones rezagadas, donde limitaciones estructurales dominan, actualmente, incentivos microeconómicos.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- Aliaga, S. (2017). Structure and financial costs for MYPES: The Peruvian case (MPRA Paper No. 91404). Munich Personal RePEc Archive. <https://mpra.ub.uni-muenchen.de/91404/>
- Banco Mundial. (2020). Doing business 2020: Comparing business regulation in 190 economies. The World Bank Group. <https://documents1.worldbank.org/curated/en/688761571934946384/pdf/Doing-Business-2020-Comparing-Business-Regulation-in-190-Economies.pdf>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Carrión-Cauja, C., Simbaña, L., & Bonilla, S. (2021). ¿El pago de impuestos genera una menor supervivencia empresarial? Un análisis de las empresas ecuatorianas de servicios. *X-Pedientes Económicos*, 5(12), 6–27. <https://portal.amelica.org/ameli/journal/392/3922449002/html/>
- Chacaltana, J. (2016). Peru, 2002-2012: Growth, structural change and formalization. *CEPAL Review*, 119, 7–23. <https://repositorio.cepal.org/server/api/core/bitstreams/df54953d-640c-499b-b8c6-19d7eea0bd99/content>
- Congreso de la República del Perú. (2013). Ley que modifica diversas leyes para facilitar la inversión, impulsar el desarrollo productivo y el crecimiento empresarial [Ley N.º 30056]. <https://cdn.www.gob.pe/uploads/document/file/3017949/Ley%2030056.pdf>
- Hemmert, G. A. J., Schons, L. M., Wieseke, J., & Schimmelpfennig, H. (2016). Log-likelihood-based pseudo-R<sup>2</sup> in logistic regression: Deriving sample-sensitive benchmarks. *Sociological Methods & Research*, 49(3), 699–728. <https://doi.org/10.1177/0049124116638107>
- Instituto Nacional de Estadística e Informática. (2020). Perú: Acceso y uso de las Tecnologías de Información y Comunicación en los hogares y por la población, 2010-2021. <https://cdn.www.gob.pe/uploads/document/file/4213770/Resumen%3A%20Per%C3%BA%3A%20Acceso%20y%20uso%20de%20las%20Tecnolog%C3%ADas%20de%20Informaci%C3%B3n%20y%20Comunicaci%C3%B3n%20en%20los%20hogares%20y%20por%20la%20poblaci%C3%B3n%2C%202010-2021.pdf?v=1677854835>
- Instituto Nacional de Estadística e Informática. (2022). V Censo Económico Nacional [Base de datos]. <https://proyectos.inei.gob.pe/microdatos/>
- Instituto Nacional de Estadística e Informática. (2025). Demografía empresarial en el Perú: IV trimestre de 2024. <https://m.inei.gob.pe/media/MenuRecursivo/boletines/boletin-demografia->



empresarial-4t24.pdf

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50(3), 649–670. <https://doi.org/10.2307/1912606>

León Mendoza, J. C., & Valcárcel Pineda, P. (2022). Influencia de las características sociodemográficas personales en el éxito empresarial en Perú. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 33, 326–352. <https://doi.org/10.46661/revmetodoscuanteconempresa.5531>

Liedholm, C. (2002). Small firm dynamics: Evidence from Africa and Latin America. *Small Business Economics*, 18(1–3), 227–242. <https://doi.org/10.1023/A:1015147826035>

Ministerio de Trabajo y Promoción del Empleo. (2024). Informe trimestral del mercado laboral: Situación del empleo en 2024, Trimestre I. <https://www.gob.pe/institucion/mtpe/informes-publicaciones/5783668-informe-trimestral-del-mercado-laboral-situacion-del-empleo-en-2024-trimestre-i>

Solomon, O. H., Allen, T., & Wangombe, W. (2024). Analysing the factors that influence social media adoption among SMEs in developing countries. *Journal of International Entrepreneurship*, 22(2), 248–267. <https://doi.org/10.1007/s10843-023-00330-9>

StataCorp. (2021). *Stata* (Version 17) [Computer software]. StataCorp LLC. <https://www.stata.com>

Tonetto, J. L., Pique, J. M., Fochezatto, A., & Rapetti, C. (2024). Survival analysis of small business during COVID-19 pandemic, a Brazilian case study. *Economies*, 12(7), 184. <https://doi.org/10.3390/economies12070184>

Torres, G. (2025). ¿Cuánto cuesta formalizar una empresa en el Perú? Trámites, requisitos, pasos y qué beneficios ofrece. *Gestión*. <https://gestion.pe/peru/cuanto-cuesta-formalizar-una-empresa-en-el-peru-tramites-requisitos-pasos-y-que-beneficios-ofrece-noticia>

Varona, L., & Gonzales, J. R. (2021). Dynamics of the impact of COVID-19 on the economic activity of Peru. *PLOS ONE*, 16(1), Article e0244920. <https://doi.org/10.1371/journal.pone.0244920>

Yamada, G. (2009). Desempeño de la microempresa familiar en el Perú. *Apuntes*, 64, 5–29. <https://www.redalyc.org/articulo.oa?id=684077011001>