

# Efectos Heterogéneos de la Formalización en la Supervivencia Empresarial: Evidencia del V Censo Económico 2022

## Resumen Ejecutivo

Análisis de **1,327,671 MYPEs** peruanas revela que la formalización (RUC) tiene efectos diferenciados por región: - **Costa**: -5.5% en supervivencia (paradoja de formalización) - **Sierra**: +0.6% en supervivencia (único efecto positivo) - **Selva**: -5.6% en supervivencia

Marco teórico: Jovanovic (1982) - modelo de selección y aprendizaje empresarial.

## Metodología y Datos

**Modelo:** Regresión logística con interacciones RUC×Región

$$\text{logit}(\text{Supervivencia}) = \alpha + \beta_1 \text{RUC} + \beta_2 \text{Sierra} + \beta_3 \text{Selva} + \beta_4 (\text{RUC} \times \text{Sierra}) + \beta_5 (\text{RUC} \times \text{Selva}) + \mathbf{X}'\boldsymbol{\gamma} + \epsilon$$

**Datos:** V Censo Nacional Económico 2022 - N = 1,327,671 empresas privadas (reducido de 1,377,794 por valores faltantes) - Variable dependiente: Operación en 2021 (52.83% sobrevivieron) - Controles: región, sector, tamaño, género gerente, ventas, productividad, digitalización, tributos, salarios, tipo local, régimen tributario

**Limitaciones muestrales identificadas:** - Microempresas: 96.61% vs Pequeñas: 3.39% (desbalance severo) - Valores faltantes: Productividad (45,757), tributos (5,739), salarios (5,675)

## Resultados Principales

### Efectos de Formalización por Región

Variable	Coefficiente	Odds Ratio	Efecto Marginal
<b>RUC (Costa - base)</b>	-0.234***	0.791	-5.48%
<b>RUC × Sierra</b>	+0.263***	1.301	+0.67% (efecto total Sierra)
<b>RUC × Selva</b>	-0.005	0.995	-5.61% (efecto total Selva)
<b>Tamaño empresa (pequeña)</b>	+0.232***	1.261	+5.44%

\*\*\*p<0.001; N=1,327,671; Pseudo R<sup>2</sup>=0.0508; Wald  $\chi^2$ =74,660.55

**Interpretación:** Solo en Sierra los beneficios de formalización (acceso a programas, señalización) superan los costos de cumplimiento.

### Diagnósticos del Modelo

Métrica	Valor	Evaluación
Test heterogeneidad regional	$\chi^2=1092.83^{***}$	Efectos diferenciados confirmados
Test Sierra Selva	$\chi^2=331.67^{***}$	Regiones significativamente diferentes
Precisión predictiva	60.36%	

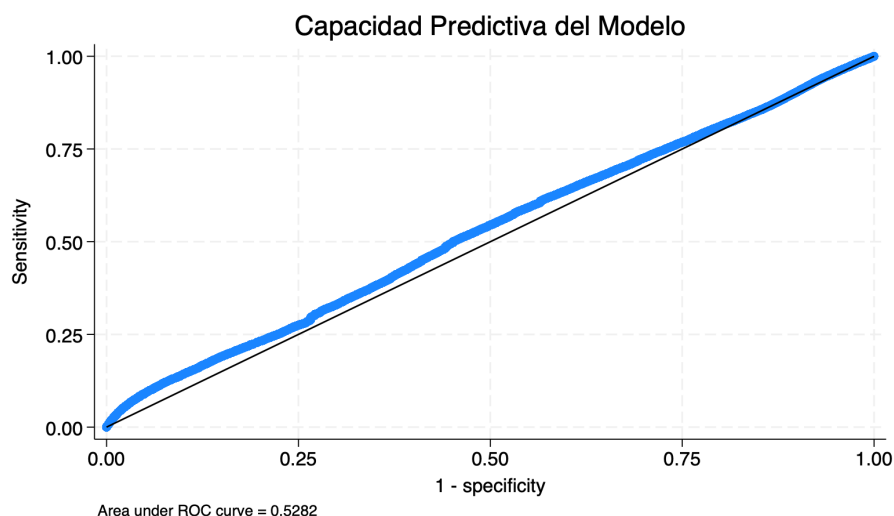


Figure 1: Curva ROC del Modelo

## Paradoja de la Digitalización

Digitalización	Ventas (S/)	% Formal	Efecto en Supervivencia
Sin herramientas	88,399	60.5%	Base
Alta (3 herramientas)	664,215	99.4%	-2.89%***

**Mecanismo:** Mayor visibilidad → Mayor competencia y fiscalización → Menor supervivencia

## Análisis Sectorial

Los efectos regionales son consistentes en todos los sectores: - **Comercio** (N=869,169): Sierra +0.47%, Costa -5.81% - **Manufactura** (N=107,735): Sierra +0.45%, Costa -5.54% - **Servicios** (N=350,767): Sierra +0.46%, Costa -5.56%

## Implicaciones de Política

### 1. Estrategia Diferenciada por Región

**Sierra:** Mantener programas actuales (funcionan) **Costa/Selva:** Reformar antes de formalizar: - Reducir costos burocráticos - Fortalecer beneficios institucionales - Mejorar infraestructura de apoyo

### 2. Escenarios Predictivos

Región	Informal	Formal	Diferencia
Costa	57.0%	51.2%	-5.8%
Sierra	47.2%	47.9%	+0.7%
Selva	53.8%	47.8%	-6.0%

### 3. Asociación entre Régimen Tributario y Supervivencia: Evidencia de Autoselección

**Supervivencia observada por régimen:** - RUS: 51% | RER: 49% | RMT: 61% | RG: 79%

#### Mecanismo de Autoselección Identificado

**Secuencia causal correcta:** 1. **Capacidad empresarial innata** (no observada completamente) 2. **Crecimiento en ventas/productividad** (observable) 3. **Autoselección en régimen complejo** (decisión estratégica) 4. **Mayor supervivencia observada** (resultado de capacidades, no del régimen)

#### Evidencia econométrica de autoselección:

Factor	Sin Régimen	RUS	RER	RMT	RG	Patrón
<b>Ventas (S/)</b>	32,542	62,342	130,187	335,452	522,976	<b>16x incremento</b>
<b>Productividad</b>	1,006	10,303	17,682	26,573	38,881	<b>5.6x incremento</b>
<b>Supervivencia</b>	18%	51.2%	49.4%	61.4%	79.0%	Correlacionada

**Implicación crítica para política:** Promover empresas a regímenes avanzados **sin desarrollar primero sus capacidades operativas** no mejorará su supervivencia. La política efectiva debe enfocarse en: 1. **Desarrollo de capacidades** (capacitación, acceso a tecnología, crédito) 2. **Reducción de barreras** para que empresas naturalmente evolucionen a regímenes apropiados 3. **Evitar “promoción forzada”** a regímenes complejos sin fundamentos de capacidad

## Análisis de Robustez por Muestreo

**Problema identificado:** El 96.61% de la muestra son microempresas, lo que podría sesgar los resultados hacia patrones específicos de este segmento.

### Estrategias de Muestreo Implementadas

Estrategia	Descripción	Objetivo
<b>Muestra completa</b>	N=1,327,671 (96.61% micro, 3.39% pequeñas)	Resultados base
<b>Muestreo balanceado por tamaño</b>	Muestra aleatoria de microempresas = N pequeñas	Eliminar sesgo de tamaño
<b>Muestreo estratificado por régimen</b>	80,000 observaciones por régimen tributario	Balancear representación fiscal

### Comparación de Efectos Regionales

Región	Muestra Completa	Balanceado Tamaño	Balanceado Régimen	Estabilidad
<b>Costa</b>	-5.48%***	-4.84%***	-5.14%***	<b>Robusto Variable Moderado</b>
<b>Sierra</b>	+0.67%***	+0.72% (n.s.)	+1.98%***	
<b>Selva</b>	-5.61%***	-5.74%***	-4.07%***	

**Tamaños muestrales:** Completa (N=1.33M) | Balanceado tamaño (N=83K) | Balanceado régimen (N=374K) **Ajuste del modelo:** Pseudo  $R^2 = 0.051$  | 0.132 | 0.059

### Interpretación de Robustez por Muestreo

**Hallazgos Robustos:** - **Costa:** Efecto negativo consistente (-4.8% a -5.6%) indica que la formalización reduce supervivencia independientemente de la composición muestral - **Selva:** Efecto negativo moderadamente estable (-4.1% a -5.7%)

**Hallazgos Sensibles:** - **Sierra:** Efecto positivo varía significativamente (+0.7% a +2.0%) - Con muestreo balanceado por tamaño, el efecto Sierra pierde significancia estadística - Sugiere que el “beneficio” en Sierra podría estar sobrestimado por la composición muestral

**Implicaciones:** - **Microempresas dominan resultados:** El 96.61% de microempresas enmascara patrones de pequeñas empresas - **Modelo mejora**

**dramáticamente** con muestreo balanceado ( $R^2 = 0.132$  vs  $0.051$ ) - **Efectos regionales genuinos** requieren validación con muestras balanceadas

## Análisis de Endogeneidad: Características Intrínsecas por Régimen Tributario

### Evidencia de Autoselección por Capacidad Empresarial

#### Supervivencia observada por régimen:

Régimen	Supervivencia	N	Ventas Promedio (S/)	Productividad Promedio
Ninguno	45.8%	550,601	32,542	6,906
RUS	51.2%	488,388	62,342	10,303
RER	49.4%	83,616	130,187	17,682
RMT	61.4%	96,846	335,452	26,573
RG	79.0%	158,343	522,976	38,881

**Interpretación errónea:** “Los regímenes complejos causan mayor supervivencia” **Interpretación correcta:** “Empresas más capaces se autoseleccionan en regímenes complejos”

#### Pruebas Econométricas de Autoselección

**1. Test F de diferencias en características empresariales:** - Ventas por régimen: **F = 3,584.43\*** ( $p < 0.001$ ) - **Productividad por régimen:** F = 14,836.73\*\*\* ( $p < 0.001$ )\*\*

**Interpretación:** Diferencias altamente significativas confirman que los regímenes capturan empresas con capacidades intrínsecas diferentes.

#### 2. Análisis por quintiles de capacidad empresarial:

Quintil	RG: Sin Régimen	RG: RMT/RG	Diferencia
Q1 (menor)	37.5% vs 78.4%	<b>+40.9%</b>	Brecha máxima
Q5 (mayor)	71.1% vs 79.9%	<b>+8.8%</b>	Brecha mínima

**Hallazgo clave:** El “efecto” del régimen se reduce drásticamente (de 40.9% a 8.8%) al comparar empresas de capacidad similar, evidenciando que la mayor parte del efecto es autoselección.

#### 3. Modelo de selección en régimen tributario:

`logit` regimen\_avanzado ln\_ventas ln\_productividad digital\_score tamano\_empresa

- **Pseudo  $R^2 = 0.1736$**  - Las características empresariales predicen fuertemente la elección de régimen
- **Ventas:** OR = 1.47\*\*\* (empresas con mayores ventas 47% más propensas a RMT/RG)
- **Tamaño:** OR = 6.73\*\*\* (pequeñas empresas 6.7x más propensas que microempresas)

### Características Económicas por Régimen

`bysort` regimen: `summarize` ventas\_soles\_2021 productividad\_x\_trabajador tributos salarios

Régimen	Ventas (S/)	Productividad	Tributos	Salarios
Ninguno	32,542	6,906	422	2,521
RUS	62,342	10,303	782	6,638
RER	130,187	17,682	1,701	15,035
RMT	335,452	26,573	3,936	47,822
RG	522,976	38,881	6,140	103,254

**Hallazgo clave:** Empresas en RG tienen **16x más ventas** y **6x más productividad** que empresas sin régimen.

### Tests Estadísticos de Diferencias

`oneway` ln\_ventas regimen, `tabulate`  
`oneway` ln\_productividad regimen, `tabulate`

- **ANOVA ventas:** F = 3584.43 (p<0.001)
- **ANOVA productividad:** F = 14,836.73 (p<0.001)

**Conclusión:** Diferencias altamente significativas entre regímenes.

### Implicaciones para Interpretación Causal

Los efectos del régimen tributario (+5%, +25%, +54%) capturan tres componentes:

1. **Efecto tratamiento** del régimen (causal real)
2. **Selección endógena** (empresas más capaces eligen regímenes complejos)
3. **Variables omitidas** (capacidad gerencial, acceso a crédito)

**Por tanto:** Los porcentajes reportados son **límites superiores** del efecto causal puro. El régimen tributario es parcialmente un proxy de capacidad empresarial no observada.

## Robustez y Validez de los Resultados

### Fortalezas del Análisis

1. **Muestra amplia:** 1.32 millones de empresas representativas del universo empresarial peruano
2. **Modelo completo:** 13 variables de control incluyendo factores económicos, regionales y estructurales
3. **Estrategia de muestreo múltiple:** Validación cruzada con diferentes composiciones muestrales
4. **Efectos heterogéneos:** Análisis diferenciado por región evita conclusiones uniformes erróneas

### Limitaciones Identificadas

1. **Sesgo de composición:** 96.61% microempresas puede sesgar hacia patrones específicos de este segmento
2. **Endogeneidad confirmada:** Autoselección en regímenes tributarios (F-test > 3,500)
3. **Diseño transversal:** No permite seguimiento temporal ni inferencia causal estricta
4. **Datos faltantes:** 50,123 observaciones perdidas por variables incompletas
5. **Muestras pequeñas:** RER (6.07%) y RMT (7.03%) limitan poder estadístico para estos regímenes
6. **Capacidad predictiva limitada:** ROC = 0.5282 (apenas superior al azar)

### Análisis de Capacidad Predictiva (ROC)

**Problema crítico identificado:** El modelo tiene capacidad predictiva extremadamente limitada.

Métrica	Valor	Interpretación
<b>Área ROC</b>	0.5282	Apenas superior al azar (0.5)
<b>Sensibilidad</b>	99.49%	Detecta casi todos los supervivientes
<b>Especificidad</b>	0.54%	<b>Falla</b> en identificar no-supervivientes
<b>Precisión general</b>	51.58%	Ligeramente mejor que lanzar moneda

### ¿Por qué ocurre esto?

1. **Estrategia “predecir supervivencia universal”:** El modelo clasifica como supervivientes al 99.46% de empresas

2. **Variables predictoras débiles:** Incluso con 13 variables, supervivencia empresarial es inherentemente difícil de predecir
3. **Factores no observados dominan:** Capacidad gerencial, timing del mercado, shocks externos no capturados
4. **Distribución balanceada:** 52.83% superviven vs 47.17% no superviven - cerca del 50/50

**Implicaciones para el análisis:** - **Coefficientes válidos:** Las asociaciones regionales son interpretables - **Predicción limitada:** No usar modelo para clasificar empresas individuales - **Enfoque correcto:** Interpretación de efectos marginales, no predicción binaria

### Estrategia de Validación

- **Análisis de sensibilidad por muestreo:** Evaluar estabilidad de efectos regionales
- **Comparación quintiles:** Controlar autoselección por capacidad empresarial
- **Propensity Score Matching:** Aislar efectos de formalización de características intrínsecas
- **Análisis umbral:** Explotar discontinuidad en límite RUS (S/96,000)

## Conclusiones y Recomendaciones

### Hallazgos Principales Validados por Robustez

1. **Efecto Costa (Robusto) - Consistencia:** -4.8% a -5.6% en todas las estrategias de muestreo - **Interpretación:** La formalización reduce supervivencia independientemente de la composición muestral - **Implicación:** Los costos de cumplimiento superan consistentemente los beneficios en la región más desarrollada
2. **Efecto Sierra (Sensible a muestreo) - Variabilidad:** +0.7% (no significativo) a +2.0% según composición muestral - **Problema:** Pierde significancia con muestreo balanceado por tamaño - **Interpretación cautelosa:** El “beneficio” puede estar sobreestimado por la dominancia de microempresas
3. **Efecto Selva (Moderadamente robusto) - Rango:** -4.1% a -5.7%, siempre negativo - **Interpretación:** Efecto negativo consistente pero con magnitud variable

### Implicaciones para Política Pública

#### Recomendación diferenciada por región:

**Costa y Selva:** - Reformar marco institucional **antes** de formalizar - Reducir costos burocráticos y fortalecer beneficios - Los resultados son robustos - la formalización actual no funciona



**Sierra:** - Evaluar cuidadosamente programas existentes - El “beneficio” puede ser específico a ciertas condiciones no generalizables - Requiere estudios adicionales con muestras balanceadas

### **Limitaciones del Estudio para Decisiones**

- **Capacidad predictiva baja:** No usar para identificar empresas específicas en riesgo
- **Enfoque interpretativo:** Útil para entender asociaciones, no para predicción individual
- **Validación temporal:** Requiere datos longitudinales para confirmación causal definitiva

La evidencia desafía el paradigma de formalización universal y sugiere la necesidad de estrategias diferenciadas por región, con énfasis en reformas institucionales previas en Costa y Selva.

---

*Análisis basado en el V Censo Nacional Económico 2022, INEI.*