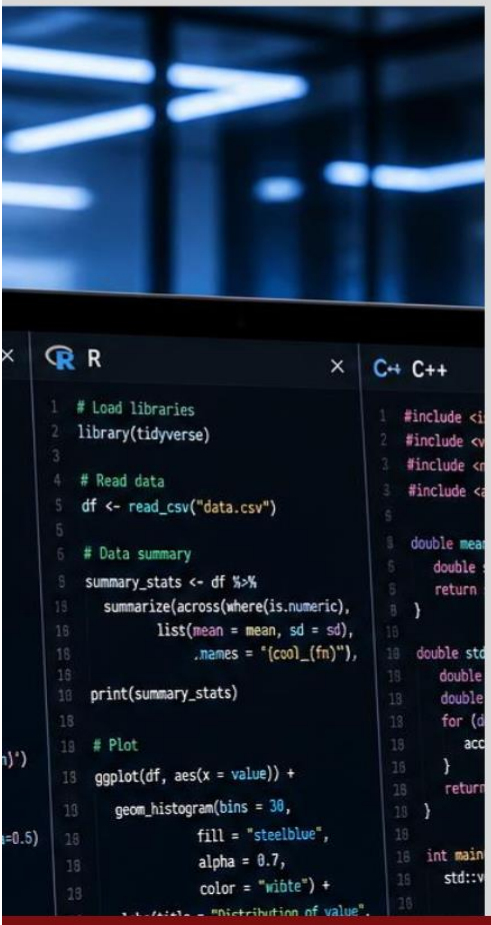
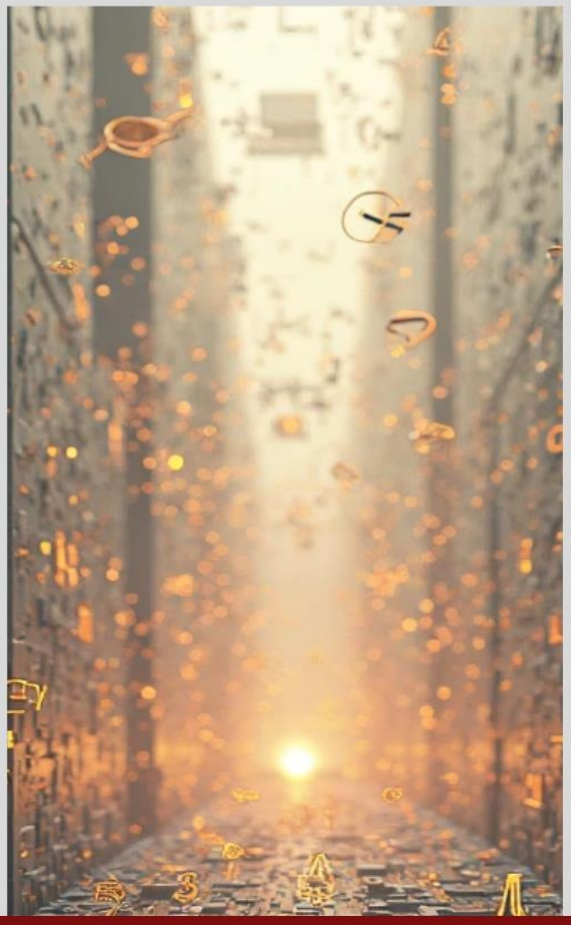
AUSTRAL





MODELOS ECONOMÉTRICOS Y ESTADÍSTICOS BASADOS EN DATOS DE LA SUNAT, EL BCRP Y EL INEI

Carolina Montes Lizárraga, Raúl Yajahuanca Huancas, Ana Cristhina Blanco Napuri De Martínez, Evelyn Esteffany Ordinola Lama, Edwin Johny Asnate Salazar, Bebeto Juan Pizarro Marcos

ISBN: 978-9915-720-05-0

9 789915 720050

WWW.EDITORIALAUSTRAL.COM

Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCRP y el INEI

Montes Lizárraga, Carolina; Yajahuanca Huancas, Raúl; Blanco Napuri De Martínez, Ana Cristhina; Ordinola Lama, Evelyn Esteffany; Asnate Salazar, Edwin Johny; Pizarro Marcos, Bebeto Juan

© Montes Lizárraga, Carolina; Yajahuanca Huancas, Raúl; Blanco Napuri De Martínez, Ana Cristhina; Ordinola Lama, Evelyn Esteffany; Asnate Salazar, Edwin Johny; Pizarro Marcos, Bebeto Juan, 2026

Coordinador de edición y volumen:

Andrea Mercedes Alvarez Rubio, 2026

Primera edición (1.ª ed.): septiembre, 2026

Editado por:

Editorial Austral SRL ®

www.editorialaustral.com

Puntas de Santiago 1694, 11500, Montevideo, Departamento de Montevideo, Uruguay.

Diseño de carátula e ilustraciones: Editorial Austral SRL ®

Libro electrónico disponible en:

https://purl.archive.org/purl/austral/9789915720050

Formato: Electrónico

ISBN: 978-9915-720-05-0

Código de conducta OASPA: Apoyamos el acceso abierto de acuerdo con el código de conducta, la transparencia y las mejores prácticas de OASPA para la publicación de libros académicos y de investigación. Estamos comprometidos con los más altos estándares editoriales en ética y deontología, bajo la premisa de «Historias con la fuerza del ciento austral»

Los lineamientos para la revisión por pares a doble ciego de Editorial Austral SRL definen normas y principios esenciales que todos los revisores tienen que acatar a lo largo del procedimiento de revisión de libros. Estas normas son pertinentes para varias disciplinas y pueden servir como guía a sus autores; además, son una herramienta pedagógica que utiliza la editorial para formar su comité, de acuerdo con las Directrices éticas para revisores por pares de COPE.

CC BY-NC 4.0: Esta obra se puede compartir, copiar y redistribuir en cualquier medio o formato. Adaptar, remezclar, transformar y construir a partir del material. No se puede hacer uso del material con propósitos comerciales.

Editorial Austral se adhiere a la "Recomendación relativa a la preservación del patrimonio documental, comprendido el patrimonio digital, y el acceso al mismo" de la UNESCO y a la Norma Internacional de referencia para un sistema abierto de información archivística (OAIS-ISO 14721). Este libro está preservado digitalmente por datasegura.info

Editorial Austral

Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCRP y el INEI

Montevideo, 22 de septiembre de 2026

Índice

Resumen 7

Abstract 8

Prólogo 9

Introducción 11

Capítulo 1: Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCP y el INEI 15

- Arquitectura institucional y la naturaleza multifrecuencia de los datos económicos 16
- Metodologías Econométricas del INEI: cuentas nacionales y microeconometría social 17
- Estimación del PBI, descomposición temporal y agregación macro 17
- Microeconometría aplicada a la ENAHO: modelación de pobreza e ingresos 18
- Analítica predictiva y modelos estadísticos en la SUNAT: estimación de brechas y control de riesgos 20
- Estimación de brechas e incumplimiento tributario mediante métodos macro y series de tiempo 21
- Machine learning y detección inteligente del incumplimiento fiscal en tiempo real 22
- Modelos de alta frecuencia y nowcasting del BCP: integración de big data y proyecciones 23
- El Índice big data de consumo y la monitorización transaccional ... 23
- Modelos econométricos de nowcasting y algoritmos de aprendizaje automático 24
- Síntesis causal y triangulación de información macroeconómica ... 26

Capítulo 2: Modelos de pronóstico inmediato del producto bruto interno: fundamentos teóricos, desarrollos metodológicos y evidencia empírica en bancos centrales 30

- Arquitectura econométrica tradicional: factores dinámicos, ecuaciones puente, MIDAS y VAR de frecuencia mixta 31

La frontera del aprendizaje automático (machine learning) y el procesamiento de datos no estructurados 36
Análisis empírico comparativo y aplicaciones institucionales en bancos centrales 40
Implicancias operativas, transparencia e integración en la política monetaria 45
Capítulo 3: Modelos de proyección de recaudación tributaria: integración del PBI nominal, deflactores sectoriales y elasticidad fiscal 49
Fundamentos teóricos: elasticidad tributaria versus flotabilidad (Boyanza) 50
La descomposición nominal vs. real: el impacto de la inflación y los deflactores 52
Deflactores sectoriales y desagregación por ramas de actividad 54
Arquitectura econométrica para la proyección macrofiscal 56
Modelos de cointegración y corrección de errores (VECM y ARDL) 57
Asimetrías cíclicas y comportamiento no lineal 58
Modelación univariada y alta frecuencia (SARIMA) 58
Capítulo 4: Modelos de elasticidad y comportamiento del consumo privado: microdatos de la Encuesta Nacional de Hogares (ENAHO) y variaciones del Índice de Precios al Consumidor (IPC) 63
Marco teórico y especificación de sistemas integrados de demanda 63
Evolución de los modelos econométricos de demanda 64
Metodología econométrica para microdatos de hogares e indicadores de precios 66
Construcción de precios relativos: valores unitarios y deflactación espacial 66
Corrección por sesgo de selección y consumo nulo 67
Estimación empírica de elasticidades en la economía peruana 68
Heterogeneidad por quintiles de ingreso y sustitución de bienes 69

Impacto distributivo de la inflación y evaluación de bienestar 70
Métricas de bienestar: variación compensada 71
Dinámica de ajuste intertemporal: oferta laboral y flexibilidad del ingreso 73
Implicancias de política económica y protección social 74
Eficiencia de transferencias monetarias frente a subsidios indirectos 74
Diseños de Impuestos Selectivos al Consumo (ISC) 75
Transmisión monetaria y anclaje de expectativas 75
Capítulo 5: Modelos de riesgo crediticio y estimación de la probabilidad de incumplimiento en el Banco de Crédito del Perú 77
Arquitectura de modelamiento crediticio y factores determinantes de la PD 78
Segmentación de portafolio, variables transaccionales y scoring dinámico 80
Marco regulatorio de la SBS y comparativa con el estándar internacional IFRS 9 83
Categorización regulatoria de deudores según la SBS 83
Determinación de pérdidas esperadas bajo IFRS 9 86
Capítulo 6: Evaluación econométrica del impacto de la facturación electrónica de la SUNAT: Modelos, variables y heterogeneidad sectorial 90
Estrategia empírica y especificación del modelo econométrico 90
Modelo de diferencias en diferencias escalonado con efectos fijos 91
Modelo de mínimos cuadrados en dos etapas (2SLS / variable instrumental) 92
Operacionalización de variables y mecanismos de atenuación fiscal 93
Mecanismo de transmisión y la fricción del crédito fiscal acumulado 95

Factores explicativos de la heterogeneidad sectorial 98
Efectos de desbordamiento en redes interempresariales (*Spillovers*) 99
Hallazgos empíricos consolidados y robustez econométrica 100
Conclusión 103
Bibliografía 107

Resumen

Este libro de investigación aborda la necesidad de superar las limitaciones de los agregados macroeconómicos tradicionales y los desfases temporales en el análisis cuantitativo de la economía peruana, caracterizada por su alta informalidad y heterogeneidad estructural. Con el propósito de desarrollar, estimar y validar un conjunto de modelos econométricos y estadísticos avanzados utilizando microdatos y datos agregados provenientes de la SUNAT, el BCP y el INEI, se empleó una metodología cuantitativa e híbrida que integró microdatos de facturación y cumplimiento fiscal de la SUNAT, métricas de transaccionalidad, liquidez y comportamiento crediticio del BCP, e indicadores oficiales de PBI, inflación y empleo del INEI, procesados mediante modelos econométricos de series de tiempo (VAR, VEC, ARIMA), datos de panel y algoritmos de *machine learning* como XGBoost y Random Forest. Los resultados evidencian que la facturación electrónica y la transaccionalidad bancaria funcionan como indicadores adelantados (*leading indicators*) altamente efectivos que anteceden entre 30 y 60 días a las cifras oficiales de consumo e inversión del INEI, reduciendo significativamente el error cuadrático medio (RMSE) en las estimaciones del PBI trimestral y mejorando la identificación de brechas de informalidad regional. Se concluye que la articulación analítica de fuentes tributarias, financieras y estadísticas oficiales proporciona una arquitectura predictiva superior y de alta frecuencia, indispensable para fortalecer el diseño de políticas públicas, la gestión del riesgo crediticio y la toma de decisiones estratégicas en los sectores público y privado del Perú.

Palabras clave: toma de decisiones, sensibilidad macromeconómica, salud financiera, analítica predictiva, microeconometría social

Abstract

This research book addresses the need to overcome the limitations of traditional macroeconomic aggregates and time lags in the quantitative analysis of the Peruvian economy, which is characterized by a high degree of informality and structural heterogeneity. To develop, estimating, and validating a set of advanced econometric and statistical models using microdata and aggregate data from SUNAT, the BCP, and the INEI, a quantitative and hybrid methodology was employed that integrated microdata on revenue and tax compliance from SUNAT, metrics on transaction volume, liquidity, and credit behavior from the BCP, and official indicators of GDP, inflation, and employment from INEI, processed using econometric time-series models (VAR, VEC, ARIMA), panel data, and machine learning algorithms such as XGBoost and Random Forest. The results show that electronic invoicing and banking transactions function as highly effective leading indicators that precede between 30 and 60 days the official consumption and investment figures of INEI, significantly reducing the root mean square error (RMSE) in quarterly PBI estimates and improving the identification of gaps in regional informality. It is concluded that the analytical articulation of tax, financial, and official statistical sources provides a superior and high-frequency predictive architecture, essential to strengthen the design of public policies, credit risk management, and strategic decision-making in the public and private sectors of Peru.

Keywords: decision making, macroeconomic sensibility, financial health, predictive analytics, social microeconometrics

Prólogo

Estamos en la era del big data y la transformación digital; el análisis económico aplicado ha dejado de ser un ejercicio puramente teórico para convertirse en una disciplina impulsada por la evidencia y el rigor empírico. En el contexto peruano, articular de manera coherente el flujo de información proveniente de las principales instituciones rectoras de las cifras nacionales representa tanto un desafío metodológico como una oportunidad sin precedentes. Este libro, "Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCRP y el INEI", nace precisamente con la vocación de acortar la distancia entre la teoría cuantitativa y la realidad socioeconómica del Perú.

Las tres instituciones que sirven de pilar para esta obra constituyen el corazón del ecosistema de datos del país:

- La Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) aporta la microevidencia transaccional, el comportamiento del contribuyente, la dinámica del comercio exterior y la granularidad empresarial en tiempo casi real.

- El Banco Central de Reserva del Perú (BCRP) provee las variables macrofinancieras, las agregaciones monetarias, las tasas de interés y las expectativas del mercado indispensables para entender el ciclo económico y la estabilidad de precios.

- El Instituto Nacional de Estadística e Informática (INEI) suministra el marco estructural de la contabilidad nacional, la demografía, el empleo, los índices de precios al consumidor y los indicadores sociales provenientes de encuestas de hogares como la ENAHO.

En el texto escrito, el lector encontrará un abordaje integral que va desde el tratamiento técnico de las bases de datos (limpieza, imputación y concatenación de series temporales y de panel) hasta la especificación e interpretación de modelos econométricos avanzados. La obra cubre tanto métodos clásicos de regresión y series de tiempo (ARIMA, VAR, cointegración) como técnicas estadísticas multivariadas y enfoques contemporáneos de machine learning aplicados a la predicción de variables macroeconómicas y fiscales.

Esta edición no pretende limitarse al terreno académico; está concebida como una guía de trabajo práctica y rigurosa para investigadores, economistas, analistas de políticas públicas y estudiantes de posgrado. Al ofrecer especificaciones econométricas fundamentadas en la estructura de los datos del BCRP, el INEI y la SUNAT, entregamos una herramienta clave para formular diagnósticos certeros, realizar proyecciones de impacto e informar la toma de decisiones estratégicas tanto en el sector público como en el privado.

Confiamos en que esta obra se convertirá en un referente obligado para quienes buscan entender la economía peruana no a través de intuiciones aisladas, sino mediante la solidez del dato empírico y el análisis estadístico de alto nivel.

Montes Lizárraga, Carolina

Introducción

En el análisis económico moderno, la precisión y la oportunidad de la información son fundamentales para comprender las dinámicas de un mercado dinámico y complejo. El Perú atraviesa una etapa crucial donde el diseño de políticas públicas, la planificación financiera empresarial y la toma de decisiones estratégicas exigen herramientas cuantitativas rigurosas y fundamentadas en evidencia empírica local.

Tradicionalmente, los modelos cuantitativos han dependido de agregados macroeconómicos con desfases temporales significativas o de supuestos teóricos alejados de las particularidades de la economía peruana, caracterizada por sus niveles de informalidad, heterogeneidad productiva y bancarización en transformación. Este libro de investigación, titulado *"Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCP y el INEI"*, surge ante la necesidad de integrar tres de las fuentes primarias más robustas del país para ofrecer un marco metodológico moderno, práctico y riguroso.

Al articular la información fiscal de la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT), la información del sistema financiero privado provista por el Banco de Crédito del Perú (BCP) y los indicadores macroeconómicos y socioeconómicos oficiales del Instituto Nacional de Estadística e Informática (INEI), este trabajo conecta el microanálisis transaccional y monetario con el macroanálisis del entorno económico nacional.

Frente al reto de integrar fuentes heterogéneas para el diagnóstico y la predicción económica, la pregunta central que guía esta investigación es: ¿De qué manera la integración de datos

tributarios (SUNAT), financieros (BCP) y macroeconómicos (INEI) mediante modelos econométricos y estadísticos avanzados permite capturar la heterogeneidad de la economía peruana, optimizar la precisión de los pronósticos y mejorar la toma de decisiones públicas y privadas?

A partir de esta interrogante central, el libro desglosa las siguientes subpreguntas de análisis: ¿Cómo influyen la facturación electrónica, las declaraciones de ventas y los patrones de evasión o informalidad detectados en la SUNAT en la estimación de la actividad económica sectorial y regional?, ¿Hasta qué punto la liquidez, el comportamiento crediticio y la transaccionalidad bancaria del BCP funcionan como indicadores adelantados (*leading indicators*) del consumo privado y la inversión?, ¿Cómo interactúan los agregados nacionales (IPC, PBI, empleo) con las variables microfinancieras y tributarias al aplicar técnicas econométricas de series de tiempo, datos de panel y algoritmos de *machine learning*?

Por lo que el objetivo de investigación es desarrollar, estimar y validar un conjunto de modelos econométricos y estadísticos avanzados utilizando microdatos y datos agregados provenientes de la SUNAT, el BCP y el INEI, con el fin de proporcionar herramientas cuantitativas de alta precisión para el análisis económico, la predicción de variables clave y el soporte en la toma de decisiones en el Perú.

La justificación de este libro descansa sobre tres pilares esenciales:

- Históricamente, las fuentes tributarias, financieras y estadísticas han sido analizadas de forma aislada. La SUNAT analiza el cumplimiento fiscal; el BCP evalúa el riesgo financiero

y la liquidez; el INEI calcula la contabilidad nacional. Integrar conceptual y cuantitativamente estas tres perspectivas permite corregir sesgos de información, reducir el desfase temporal de las estadísticas oficiales y abordar fenómenos complejos como la economía informal mediante enfoques híbridos.

- Para el sector privado —especialmente el sector financiero y corporativo—, la capacidad de predecir la demanda y evaluar el riesgo de crédito mediante variables transaccionales y fiscales integradas representa una ventaja competitiva determinante. Para el sector público, la modelación combinada mejora la precisión en la proyección de la recaudación fiscal y la focalización de políticas de reactivación económica.
- Esta obra cierra una brecha en la literatura económica peruana, ofreciendo a investigadores, docentes y estudiantes de grado y posgrado un manual aplicado de econometría y estadística con datos del contexto real. Asimismo, contribuye a la transparencia y al fortalecimiento de las capacidades analíticas del país.

A lo largo de la sistematización, se abordan técnicas de econometría clásica (modelos VAR, VEC, ARIMA, modelos de datos de panel estáticos y dinámicos) junto con métodos estadísticos contemporáneos (modelos de clasificación, regresión penalizada LASSO/Ridge y algoritmos de clasificación/agrupamiento aplicados a finanzas y economía). Y el ámbito de aplicación comprende el territorio peruano a nivel nacional y descentralizado (por departamentos o regiones, según la disponibilidad de información desagregada). El horizonte de datos cubre el período comprendido entre los años 2010 y 2025/2026, lo que permite capturar ciclos expansivos, crisis exógenas y la posterior recuperación del mercado interno.

En síntesis, la investigación se enfoca en la formulación, estimación e interpretación práctica de modelos cuantitativos. La información del BCP se utiliza a nivel de agregados macrofinancieros e indicadores autorizados o anonimizados para garantizar la reserva bancaria y la protección de datos personales. Las variables de la SUNAT e INEI corresponden a datos de acceso público e informes estadísticos oficiales

Capítulo 1: Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCP y el INEI

La economía peruana se caracteriza por una marcada heterogeneidad estructural, un elevado nivel de informalidad laboral y empresarial, y una vulnerabilidad constante a choques exógenos y de oferta. Para capturar esta dinámica, formular políticas públicas efectivas y tomar decisiones de inversión en tiempo real, las instituciones públicas y los agentes del sistema financiero han desarrollado una avanzada infraestructura cuantitativa. Esta infraestructura combina la econometría tradicional de series de tiempo y modelos microeconométricos con arquitecturas modernas de *Machine Learning* (ML) y *Big Data*.

El marco analítico nacional se sostiene fundamentalmente en tres pilares informativos: el Instituto Nacional de Estadística e Informática (INEI), que estructura las cuentas nacionales y las encuestas socioeconómicas de hogares; la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT), que provee información masiva a partir de registros tributarios y Comprobantes de Pago Electrónicos (CPE); y el Área de Estudios Económicos del Banco de Crédito del Perú (BCP), que procesa información transaccional de alta frecuencia para proyectar el consumo privado y el Producto Bruto Interno (PBI) en tiempo real (Instituto Nacional de Estadística e Informática, 1999).

Arquitectura institucional y la naturaleza multifrecuencia de los datos económicos

La combinación cuantitativa de las bases de datos generadas por el INEI, la SUNAT y el BCP enfrenta la barrera de la disparidad en las frecuencias de muestreo, la latencia en la disponibilidad pública y los niveles de agregación. El ecosistema de información abarca desde encuestas estructurales con periodicidad anual hasta registros transaccionales procesados de manera continua (véase la tabla 1).

Tabla 1: Ecosistema de información económica del INEI, SUNAT y BCP

Institución	Tipo de Información Dominante	Frecuencia de Actualización	Rezago Típico de Publicación	Enfoque Principal de Modelación
INEI	Cuentas Nacionales, ENAHO, IPC, Censos Estadísticos	Mensual / Trimestral / Anual	45 a 70 días	Descomposición de series de tiempo, regresiones probabilísticas (Logit/Probit), ecuaciones de Mincer.
SUNAT	Registros de IGV/Renta, Comprobantes Electrónicos (CPE), Importaciones	Diario / Mensual / Anual	Tiempo real a 30 días	Estimación *top-down* de brechas, modelos SARIMA, clasificadores supervisados de riesgo fiscal.
BCP	Transacciones con tarjeta, transferencias (Yape), colocaciones crediticias	Diario / Semanal / Mensual	Inmediato (Alta Frecuencia)	*Nowcasting*, modelos MIDAS, regresiones penalizadas (Elastic Net), algoritmos de ensamble (GBM).

Esta discrepancia temporal genera en la econometría aplicada el fenómeno de la "información desequilibrada" (*unbalanced datasets* o *ragged edges*). Mientras el INEI reporta el PBI trimestral con un retraso promedio de 45 días respecto al cierre del periodo, la SUNAT recibe millones de registros diarios de comprobantes de pago electrónicos, y el BCP monitorea los flujos de consumo privado a través de sus canales transaccionales con inmediatez prácticamente absoluta. Para subsanar estos desfases, las tres instituciones han impulsado la transición desde modelos macroeconómicos lineales tradicionales hacia esquemas de *Nowcasting* e inteligencia artificial predictiva (INEI, 2002).

Metodologías Econométricas del INEI: cuentas nacionales y microeconometría social

El INEI constituye la columna vertebral del sistema estadístico nacional, al elaborar los agregados macroeconómicos oficiales a precios constantes y corrientes, además de ejecutar la Encuesta Nacional de Hogares (ENAHO). Los modelos aplicados sobre sus datos abordan tanto la evolución del ciclo económico como los determinantes de la pobreza y la estructura del mercado laboral.

Estimación del PBI, descomposición temporal y agregación macro

El INEI calcula el PBI mediante los enfoques de la producción, el gasto y el ingreso, utilizando como base el año 2007. Dado que el dato oficial del PBI se publica con un retraso estructural, los modelos econométricos aplicados a sus series temporales emplean técnicas de descomposición para aislar la tendencia, el ciclo y la estacionalidad.

Para la proyección y el análisis del componente cíclico del PBI, se recurre a la especificación de componentes no observados y a la estimación de variables sintéticas. La estructura de estimación directa del PBI mensual (Y_t) en función de sus componentes estacionales (Est_t) y de tendencia-ciclo (TC_t) se define mediante modelos de regresión con variables *dummy* estacionales y polinomios de tendencia (Weller, 2014):

$$\text{TC}_t = \beta_0 + \sum_{q=1}^{Q} \beta_q \; t^q + \epsilon_t$$

$$\text{Est}_t = \sum_{m=1}^{12} \gamma_m \; d_{m,t} + \vartheta_t$$

donde $d_{m,t}$ representa las variables dicotómicas para cada mes del año y t la tendencia temporal. Ante la presencia de cambios estructurales en la economía peruana —tales como choques climáticos o fluctuaciones en la producción minera y pesquera— el INEI complementa la agregación sectorial con modelos autorregresivos integrados de media móvil (ARIMA) para ajustar las variaciones estacionales mediante el procedimiento X-13ARIMA-SEATS.

Microeconometría aplicada a la ENAHO: modelación de pobreza e ingresos

A nivel microeconómico, los datos de la ENAHO permiten construir modelos probabilísticos para medir la vulnerabilidad socioeconómica y las brechas salariales en el mercado laboral peruano. En el análisis de la pobreza monetaria —que afectó al 25.7% de la población nacional en 2025 y a un 4.7% en condición de pobreza extrema— se emplean modelos de respuesta dicotómica Logit y Probit

para estimar la probabilidad de que un hogar caiga bajo la línea de pobreza (L_p).

Definiendo Y_i^* como una variable latente no observada que representa la insuficiencia del gasto per cápita del hogar i respecto a la canasta básica de consumo (valorizada en S/ 462 mensuales por persona en 2025), el modelo estadístico se especifica como:

$$Y_i^* = \mathbf{X}_i'\boldsymbol{\beta} + u_i$$

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* \leq L_p \quad \text{(Hogar en situación de pobreza)} \\ 0 & \text{si } Y_i^* > L_p \quad \text{(Hogar no pobre)} \end{cases}$$

donde $\mathbf{X}_i$ es un vector de características socioeconómicas del jefe de hogar (nivel educativo, lengua materna, acceso a servicios públicos y área de residencia urbana o rural), y u_i sigue una distribución logística o normal estándar.

De forma paralela, para el estudio de la distribución de ingresos y la estimación del potencial tributario en las rentas del trabajo, los microdatos de la ENAHO se someten a especificaciones de la ecuación de ingresos de Mincer (INEI, 2017):

$$\ln(w_i) = \alpha + \beta_1 \text{Educ}_i + \beta_2 \text{Exp}_i + \beta_3 \text{Exp}_i^2 + \mathbf{Z}_i'\boldsymbol{\gamma} + \epsilon_i$$

donde $\ln(w_i)$ representa el logaritmo natural del ingreso por hora, Educ_i los años de escolaridad formal, Exp_i la experiencia laboral potencial (y su término cuadrático para capturar rendimientos decrecientes), y $\mathbf{Z}_i$ incluye variables de control como informalidad laboral, sector económico y categoría ocupacional. La corrección por sesgo de selección muestral de Heckman se aplica comúnmente cuando se modela la decisión de participar en el mercado laboral formal versus el informal. (véase la tabla 2).

Tabla 2: Aplicación de modelos econométricos para los datos recolectados del INEI

Modelo	Fuente de Datos	Variables Predictoras Clave	Variable Objetivo	Aplicación Principal
Logit / Probit de Pobreza	ENAHO - INEI	Nivel educativo, miembros del hogar, acceso a saneamiento, región geográfica.	Estado de Pobreza Monetaria (0 o 1).	Estimación de determinantes del bienestar y efectividad de programas sociales.
Ecuación de Mincer con Selección	ENAHO - INEI	Años de educación, experiencia laboral, formalidad, sector productivo.	Logaritmo del Ingreso por Hora ($\ln w_i$).	Medición de retornos de la educación y cuantificación de la brecha de ingresos laborales.
Imputación (Hot Deck / Regresión)	ENAHO - INEI	Atributos demográficos, consumo del hogar, ubicación geográfica.	Ingresos/Gastos no declarados o sesgados.	Corrección de no respuesta para el cálculo de desigualdad y recaudación potencial.

Analítica predictiva y modelos estadísticos en la SUNAT: estimación de brechas y control de riesgos

La SUNAT ha transformado su arquitectura analítica, pasando de un esquema de fiscalización *ex post* a un modelo de gestión de riesgos basado en la explotación intensiva de datos en tiempo real. Esta transición responde a la masificación de los Comprobantes de Pago Electrónicos (CPE) y a la integración de más de 40 fuentes de información internas y externas.

Estimación de brechas e incumplimiento tributario mediante métodos macro y series de tiempo

Para evaluar la eficiencia del sistema tributario y determinar las pérdidas por evasión, elusión y morosidad, la SUNAT aplica metodologías estructurales *top-down*. La recaudación fiscal se distingue por estar compuesta por los impuestos del país, los cuales son recaudados de forma ininterrumpida por la administración tributaria y gestionados por el gobierno; está orientada a satisfacer las necesidades del Estado y está enfocada en beneficiar a la población (Ventura et al., 2024). Está sujeta a fiscalización y regida por principios como la igualdad, la reserva de ley y el respeto a los derechos fundamentales. La recaudación potencial de IGV (IT_p) se calcula a partir de las Cuentas Nacionales del INEI mediante la construcción de la base imponible teórica (Medina, 2023):

$$\begin{aligned} IT_p = \tau \times [&\text{Consumo Final de los Hogares} \\ &+ \text{Consumo Intermedio Inafecto} \\ &+ \text{Inversión Bruta Inafecta}] \end{aligned}$$

donde τ representa la tasa nominal del impuesto (18%). El porcentaje de incumplimiento tributario (m_{IGV}) se define formalmente como:

$$m_{\text{IGV}} = \left(1 - \frac{IT_e}{IT_p}\right) \times 100$$

siendo IT_e la recaudación efectiva neta reportada en los estados de gestión de la SUNAT.

Para el pronóstico de la recaudación tributaria mensual de corto plazo, la SUNAT utiliza modelos de series de tiempo de la clase SARIMA (Modelo Autorregresivo Integrado de Media Móvil Estacional). La

especificación general del modelo $(p,d,q)\times(P,D,Q)_s$ se representa mediante (Tapia, 2025):

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t$$

donde B es el operador de retardo, $s=12$ refleja la frecuencia mensual de la recaudación, d y D indican los órdenes de diferenciación ordinaria y estacional para asegurar la estacionariedad, y ϵ_t representa un proceso de ruido blanco. Estos modelos capturan los picos estacionales de la recaudación registrados en los meses de regularización del Impuesto a la Renta (marzo-abril) y los incrementos de la actividad comercial hacia fin de año.

Machine learning y detección inteligente del incumplimiento fiscal en tiempo real

Para maximizar la recaudación neta —que en 2025 alcanzó los S/ 175,156 millones con una presión tributaria de 14.6%— la SUNAT ha integrado modelos predictivos avanzados y algoritmos de aprendizaje automático en plataformas operativas como ATENEA.

El proceso de analítica fiscal se inicia con la ingesta masiva de transacciones desde los comprobantes electrónicos, las declaraciones juradas, los registros bancarios del Impuesto a las Transacciones Financieras (ITF) y las operaciones aduaneras. Esta información alimenta la plataforma ATENEA, donde operan tres familias de modelos analíticos:

En primer lugar, los algoritmos de redización y análisis de grafos detectan esquemas de facturación apócrifa o "huachicol de facturas". La inteligencia artificial evalúa si un emisor posee la capacidad operativa mínima (activos, planilla declarada, consumo de energía) para realizar las ventas registradas. Mediante esta técnica, la SUNAT

identificó 212 empresas emitientes de comprobantes falsos y dio de baja a 1,874 contribuyentes catalogados como Sujetos Sin Capacidad Operativa.

En segundo lugar, los modelos de clasificación supervisada y detección de anomalías asignan puntajes de riesgo tributario en tiempo real. Al comparar las compras y ventas declaradas contra las tendencias del sector económico, el sistema detecta inconsistencias en el débito y crédito fiscales del IGV, emitiendo alertas preventivas a los contribuyentes para fomentar la autocorrección antes del inicio de acciones punitivas (Castillo y Lahura, 2022).

En tercer lugar, la integración de algoritmos de Procesamiento de Lenguaje Natural (NLP) automatiza la revisión de expedientes, solicitudes de devolución y jurisprudencia del Tribunal Fiscal, optimizando las fiscalizaciones en dependencias regionales.

Modelos de alta frecuencia y nowcasting del BCP: integración de big data y proyecciones

El Área de Estudios Económicos del BCP elabora estimaciones sobre el desempeño macroeconómico del país para orientar la toma de decisiones crediticias, la gestión de tesorería y las proyecciones del sistema financiero. Ante el retraso de las publicaciones oficiales, la entidad ha desarrollado un conjunto de herramientas sustentado en *Big Data* y datos transaccionales.

El Índice big data de consumo y la monitorización transaccional

Dado que el consumo privado explica más del 60% del PBI en el Perú, el BCP procesa de forma agregada y anónima millones de registros transaccionales diarios provenientes de tarjetas de débito y

crédito, retiros en cajeros automáticos, transferencias electrónicas en plataformas masivas como Yape y operaciones en puntos de venta (POS). Por lo tanto, en el año 2022, las operaciones se llevaban a cabo principalmente por los canales digitales y físicos, que suponían solamente el 45% del total de las transacciones. En el 2023, estos canales representaban un 60 % y en el primer cuatrimestre del 2024 continúan creciendo, llegando a ser el 73 % de todas las transacciones (El Comercio, 2024).

El Índice Big Data de Consumo Privado se construye mediante el filtrado de estas series transaccionales, ajustándolas por inflación, días laborables y factores estacionales. La serie resultante funciona como un indicador coincidente del gasto real de los hogares. Durante periodos de aceleración de la demanda interna (con crecimientos superiores al 5%), este índice muestra un alto grado de correlación con las cifras trimestrales presentadas posteriormente por el INEI.

Modelos econométricos de nowcasting y algoritmos de aprendizaje automático

Para convertir los datos de alta frecuencia en estimaciones puntuales del PBI trimestral en tiempo real, se aplican técnicas econométricas diseñadas para lidiar con datos no alineados temporalmente.

Los Modelos de Factores Dinámicos (DFM) condensan la información de una matriz amplia de variables x_t —que incluye exportaciones, recaudación del IGV de la SUNAT, generación eléctrica y volumen transaccional del BCP— en un número reducido de factores latentes f_t que capturan la tendencia común de la economía:

$$x_t = \mathbf{C}_0 f_t + e_t, \quad e_t \sim \mathcal{N}(0, \mathbf{R})$$

$$f_t = \mathbf{A}_1 f_{t-1} + \cdots + \mathbf{A}_p f_{t-p} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \mathbf{Q})$$

Complementariamente, las Regresiones con Muestreo de Datos Mixtos (MIDAS) relacionan directamente la variable de baja frecuencia (PBI trimestral, y_t) con indicadores de alta frecuencia (mensuales o diarios, $x_m^{(f)}$), empleando polinomios de retardos distribuidos (Luo y Yu, 2021):

$$y_t = \beta_0 + \beta_1 B\left(L^{1/f}; \boldsymbol{\theta}\right) x_t^{(f)} + \epsilon_t$$

donde la función de ponderación $B\left(L^{1/f}; \boldsymbol{\theta}\right)$ parametriza la estructura de rezagos mediante especificaciones funcionales como la distribución Beta o la función de Almon.

Cuando la cantidad de regresores potenciales excede la cantidad de observador ($p > n$), el uso de Mínimos Cuadrados Ordinarios es inviable por multicolinealidad. En este escenario, se recurre al modelo Elastic Net, el cual combina penalizaciones L_1 (LASSO) y L_2 (Ridge) para seleccionar variables predictoras y manejar grupos de indicadores correlacionados (Greenwood et al., 2020):

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\, \beta_j \right)^2 + \lambda \left[\alpha \sum_{j=1}^{p} |\, \beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2 \right] \right\}$$

donde λ es el parámetro de regularización global y $\alpha \in [0,1]$ equilibra la mezcla entre la selección estricta de variables y el suavizado de coeficientes. Adicionalmente, se incorporan algoritmos no lineales como Gradient Boosting Machines (GBM) y Support Vector Regression (SVR), los cuales destacan por su capacidad de captura de puntos de inflexión durante choques macroeconómicos imprevistos (véase la Tabla 3).

Tabla 3: Modelos estadísticos de aprendizaje automático, desempeño y puntos fuertes

Familia de Modelos	Mecanismo Algorítmico	Tratamiento de Dimensionalidad y Frecuencia	Desempeño y Puntos Fuertes
Benchmark Autorregresivo (AR/SARIMA)	Regresión lineal sobre rezagos de la propia serie.	Limitado a una sola frecuencia lineal.	Útil como piso de comparación; ineficiente ante giros drásticos del ciclo.
Modelos de Factores Dinámicos (DFM)	Extracción de componentes latentes mediante filtro de Kalman.	Maneja grandes volúmenes de series; requiere estructuración en bloques.	Elevada interpretabilidad macroeconómica de los factores extraídos.
Elastic Net / Adaptive LASSO	Minimización del error cuadrático con regularización $L_1 + L_2$.	Resuelve el problema $p > n$; realiza selección automática de covariables.	Elimina la multicolinealidad entre indicadores financieros y tributarios.
Gradient Boosting Machines (GBM)	Ensamble secuencial de árboles de decisión sobre residuos.	Modela interacciones no lineales complejas entre series de alta frecuencia.	Precisión predictiva superior en el *nowcasting* de corto plazo.

Síntesis causal y triangulación de información macroeconómica

El valor analítico más determinante en el contexto peruano proviene de la triangulación de datos entre la microeconometría del INEI, los registros tributarios de la SUNAT y los datos de alta frecuencia del BCP.

Esta interconexión se articula mediante un flujo continuo de validación entre las tres fuentes. El INEI fija el marco conceptual y los

pesos de la canasta representativa en las Cuentas Nacionales y la ENAHO. La SUNAT proporciona la verificación transaccional a través del débito fiscal del IGV y los comprobantes electrónicos, permitiendo validar si las dinámicas sectoriales reportadas en las encuestas se reflejan en la facturación efectiva. A su vez, el BCP captura la velocidad de circulación del dinero y el consumo diario mediante sus plataformas digitales y bancarias, datos que alimentan los modelos de *nowcasting* para anticipar las cifras oficiales que el INEI publicará semanas después (INEI, 2017).

De esta dinámica de integración se derivan tres mecanismos clave de validación cuantitativa:

En primer lugar, la estimación del consumo privado en tiempo real combina el Índice Big Data del BCP —que registró expansiones interanuales de 18.3% a inicios de año y de 7.7% a mediados de ejercicio— con la recaudación por ventas minoristas reportada por la SUNAT. Cuando se detecta un incremento en el gasto transaccional con tarjetas que no se refleja de forma proporcional en la recaudación del IGV de un sector, los modelos predictivos de la SUNAT identifican focos potenciales de subdeclaración u omisión fiscal.

En segundo lugar, la cuantificación de la economía informal articula los microdatos de la ENAHO del INEI con el registro de contribuyentes de la SUNAT. La brecha entre el ingreso total estimado a través de las ecuaciones de Mincer en la ENAHO y las rentas declaradas ante la SUNAT permite calcular el potencial tributario no capturado en el mercado de trabajo independiente. A la par, el BCP evalúa el ritmo al que la adopción de billeteras digitales desplaza el uso de efectivo, reduciendo el margen de la informalidad no registrada.

En tercer lugar, la calibración del *nowcasting* macroeconómico utiliza un enfoque multivariado. Los modelos predictivos que combinan el PBI de sectores primarios del INEI, la recaudación aduanera por importaciones de bienes de capital de la SUNAT y la liquidez del sistema bancario procesada por el BCP reducen sustancialmente el Error Cuadrático Medio de Pronóstico (RMSE) en comparación con modelos autorregresivos aislados (Pérez, 2021).

El desarrollo econométrico y estadístico en el Perú demuestra que la combinación de fuentes de datos tradicionales con analítica avanzada de *Big Data* permite mitigar el impacto del rezago en la información oficial. Mientras el INEI garantiza la consistencia metodológica estructural de las cuentas nacionales y los indicadores de bienestar socioeconómico, la SUNAT y el BCP expanden la frontera del monitoreo en tiempo real mediante *Machine Learning* y procesamiento masivo de datos transaccionales.

A partir del análisis realizado, se establecen las siguientes recomendaciones para el perfeccionamiento de la modelación económica e institucional:

1. Implementación de Ensembles Multimodelo para Proyecciones Oficiales: Se recomienda a los equipos de investigación económica adoptar esquemas de ensamble predictivo que combinen ponderadamente las estimaciones de Modelos de Factores Dinámicos (DFM) y algoritmos de regularización (Elastic Net) con redes neuronales o algoritmos de boosting. Esta metodología minimiza el sesgo de estimación frente a choques no estructurales.

2. **Escalamiento del Análisis de Redes Fiscales en la SUNAT**: Es prioritario profundizar la integración de la teoría de grafos sobre la plataforma ATENEA para monitorear la totalidad de la red de Comprobantes de Pago Electrónicos. La identificación temprana de nodos atípicos en la emisión de facturas permite actuar de forma preventiva antes del cierre de los ejercicios fiscales.

3. **Fomento de Entornos Interoperables de Microdatos Anonimizados**: Se sugiere promover esquemas de colaboración institucional que permitan cruzar, bajo estrictos protocolos de privacidad, los paneles de hogares del INEI con los registros de bancarización del BCP y los historiales de cumplimiento de la SUNAT. La disponibilidad de estos microdatos combinados facilitará el diseño de políticas públicas de formalización e inclusión financiera basadas en evidencia causal rigurosa.

Capítulo 2: Modelos de pronóstico inmediato del producto bruto interno: fundamentos teóricos, desarrollos metodológicos y evidencia empírica en bancos centrales

El seguimiento del estado de la economía en tiempo real constituye un pilar crítico para la formulación de la política monetaria y la evaluación de la estabilidad macroeconómica. No obstante, la medición de la actividad económica agregada a través del Producto Bruto Interno (PBI) enfrenta una restricción estructural insoslayable: el rezago de publicación (publication lag). Las agencias oficiales de estadística difunden las cifras del PBI con semanas o meses de retraso respecto al cierre del trimestre o mes de referencia debido al tiempo requerido para compilar, consolidar y auditar la información proveniente de encuestas empresariales, registros tributarios y cuentas administrativas. En este lapso de opacidad, las autoridades económicas deben adoptar decisiones estratégicas en un entorno caracterizado por una marcada incertidumbre informativa (INEI, 2017).

Para atenuar este problema de información asíncrona surgió la metodología de pronóstico inmediato o nowcasting, un término adaptado de la meteorología que en la ciencia económica designa la estimación en tiempo real del estado actual de la economía, abarcando el presente, el pasado reciente no publicado y el futuro inminente. A diferencia del pronóstico macroeconómico tradicional a mediano

plazo —centrado en trayectorias estructurales de equilibrio—, el nowcasting tiene como objetivo la extracción eficiente de señales cuantitativas oportunas a partir del flujo diario, semanal y mensual de indicadores económicos parciales para predecir un agregado macroeconómico de menor frecuencia publicado con retraso.

Desde la perspectiva econométrica, el desafío fundamental del nowcasting consiste en sistematizar una matriz de datos altamente heterogénea caracterizada por dos complejidades operativas interrelacionadas: la frecuencia mixta y el problema del borde irregular (ragged-edge problem). Por un lado, las variables explicativas potenciales (como la producción de energía eléctrica, las ventas al por menor, la recaudación del impuesto general a las ventas, los índices de confianza empresarial y las importaciones) se reportan a frecuencias mensuales, semanales o diarias, mientras que la variable objetivo (PBI) se mide a frecuencia mensual o trimestral (INEI, 2002).

Por otro lado, la difusión de estos indicadores parciales ocurre de forma desincronizada a lo largo del mes, generando una matriz de información asimétrica en el extremo muestral más reciente. Resolver esta estructura asíncrona exige marcos capaces de condensar paneles masivos de datos, proyectar observaciones faltantes sin distorsionar la señal estocástica subyacente y adaptarse a rupturas estructurales en el ciclo económico.

Arquitectura econométrica tradicional: factores dinámicos, ecuaciones puente, MIDAS y VAR de frecuencia mixta

La econometría macroeconómica desarrollada en los bancos centrales articuló una primera generación de modelos orientados a

formalizar el tratamiento de frecuencias mixtas y la reducción de dimensionalidad. Entre estas metodologías destacan los Modelos de Factores Dinámicos (DFM), las Ecuaciones Puente (Bridge Models), las regresiones de Muestreo de Datos de Frecuencia Mixta (MIDAS) y los Vectores Autorregresivos de Frecuencia Mixta (MF-VAR) (Brum y Rodríguez, 2016).

Los Modelos de Factores Dinámicos (DFM), representan el estándar histórico de referencia en los institutos emisores. La premisa económica de los DFM es que las co-movilidades de un panel amplio de indicadores macroeconómicos y financieros $x_{it}(i = 1, \dots, N)$ pueden resumirse mediante un número reducido de factores latentes comunes $f_t \in \mathbb{R}^r$ $(r \ll N)$, suplementados por componentes de error idiosincrásicos e_{it} que capturan el ruido específico de cada serie (Coroneo et al., 2016). Matemáticamente, la representación general en el espacio de estados de un DFM estandarizado se formula mediante la ecuación de observación y la ecuación de transición:

$$x_{it} = \lambda_i^{\top} f_t + e_{it}, \quad e_{it} \sim \mathcal{N}(0, \psi_i)$$

$$f_t = \sum_{p=1}^{P} A_p f_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u)$$

En esta formulación, $\lambda_i$representa el vector de cargas del factor (factor loadings) y las dinámicas del vector de factores no observados $f_t$se rigen por un proceso autorregresivo multivariado impulsado por choques comunes u_t . La estimación de los DFM se realiza habitualmente mediante un algoritmo de dos etapas basado en Componentes Principales o a través de Máxima Verosimilitud exacta combinada con el Filtro de Kalman y el Suavizador de Kalman (Pelegrí, 2019).

La fortaleza distintiva del Filtro de Kalman radica en su capacidad recursiva para imputar los datos faltantes ocasionados por la publicación asíncrona al final de la muestra (ragged edge), lo que permite actualizar de forma continua el pronóstico del PBI cada vez que se publica un nuevo dato. Sin embargo, la rigidez implícita en la linealidad de los factores latentes puede restar precisión al modelo en entornos de perturbaciones atípicas o cambios estructurales bruscos (Pelegrí, 2019).

Las Ecuaciones Puente (Bridge Equations), proyectadas por Brum y Rodríguez (2016), adoptan un enfoque desagregado de carácter contable o sectorial (bottom-up). En lugar de extraer un factor latente abstracto de un panel masivo, las ecuaciones puente estiman el PBI agregando directamente proyecciones de sus componentes constitutivos (por el lado del gasto o de la oferta sectorial). Cada componente de frecuencia baja y_q se vincula linealmente con la agregación temporal de sus respectivos indicadores de alta frecuencia $x_{i,q}$mediante estimaciones de Mínimos Cuadrados Ordinarios (OLS):

$$y_q = \beta_0 + \sum_{l=1}^{m} \gamma_l \, y_{q-l} + \sum_{i=1}^{n} \beta_i \, x_{i,q} + \varepsilon_q$$

Dado que los indicadores mensuales no suelen cubrir la totalidad del período trimestral a proyectar, los valores faltantes en el horizonte de pronóstico se auxilian mediante modelos univariados de series de tiempo (típicamente especificaciones ARIMA). Una vez estimadas las ecuaciones individuales, los pronósticos de los subcomponentes se agregan aplicando las reglas contables oficiales de las cuentas nacionales, tales como ponderaciones de volumen de precios fijos o índices encadenados. Las ecuaciones puente destacan

por su transparencia analítica y por permitir una trazabilidad precisa del aporte de cada sector o componente de gasto al pronóstico general, aunque presentan una capacidad más limitada para filtrar el ruido aleatorio en matrices de datos de dimensión masiva.

Por su parte, las regresiones de Muestreo de Datos de Frecuencia Mixta (MIDAS) evitan la agregación temporal previa al enlazar directamente la variable dependiente de baja frecuencia con los rezagos de las variables explicativas de alta frecuencia. Para controlar la proliferación de parámetros que afectaría la precisión de la estimación al incluir múltiples rezagos de alta frecuencia, MIDAS impone restricciones funcionales parametrizadas sobre los coeficientes (Gordón y Camargo, 2015).

Una especificación representativa de un modelo MIDAS provisto de un polinomio restringido de Almon de segundo grado se expresa mediante la siguiente relación funcional:

$$y_q = \beta_0 + \sum_{l=1}^{m} \gamma_l \, y_{q-l} + \sum_{i=1}^{n} \sum_{h=1}^{L(i)} \delta_{i,h}\,(w) x_{i,3q-h} + \varepsilon_q$$

$$\delta_{i,h}(w_1, w_2) = \frac{\exp\left(w_{1,i} h + w_{2,i} h^2\right)}{\sum_{k=1}^{L(i)} \exp\left(w_{1,i} k + w_{2,i} k^2\right)}$$

Cuando la diferencia entre la frecuencia de los regresores y la variable objetivo no es desproporcionada —por ejemplo, al relacionar datos mensuales con agregados trimestrales—, la investigación empírica recurre con frecuencia a la variante MIDAS Sin Restricciones (U-MIDAS), introducida por Forini et al. (2011). U-MIDAS elimina el polinomio no lineal y estima los coeficientes directamente mediante OLS, simplificando el proceso de optimización computacional sin sacrificar capacidad predictiva.

En síntesis, los modelos de vectores autorregresivos de frecuencia mixta (MF-VAR), sistematizados por Schorfheide y Song (2015), capturan la dinámica conjunta endógena entre las variables trimestrales y los indicadores mensuales, expresando el sistema autorregresivo multivariado en el espacio de estados. Para evitar la sobreparametrización y la inestabilidad en la estimación derivadas de la proliferación de términos dinámicos, estos sistemas suelen estimarse bajo enfoques bayesianos (MF-BVAR) que imponen distribuciones a priori de contracción (véase la tabla 4).

Tabla 4: Dinámica conjunta endógena entre las variables operativas

Metodología Econométrica	Especificación Formal / Representación	Tratamiento del Borde Irregular (Ragged Edge)	Ventajas Analíticas y Operativas	Limitaciones Diagnósticas
Modelo de Factores Dinámicos (DFM)	Sistema en Espacio de Estados: $x_{it} = \lambda_i^{\top} f_t + e_{it}$	Algoritmo iterativo Filtro/Suavizador de Kalman	Extracción eficiente de señales en paneles masivos; reducción de ruido	Supuesto de linealidad; menor ajuste ante choques estructurales
Ecuaciones Puente (Bridge Equations)	Regresión por componentes + OLS: $y_q = \beta_0 + \sum \beta_i \, x_{i,q}$	Relleno univariado de series faltantes (ARIMA)	Transparencia contable; descomposición sectorial directa	Susceptible a la volatilidad de indicadores individuales
Regresiones MIDAS / U-MIDAS	Polinomios de Almon sobre rezagos de alta frecuencia	Preservación de la frecuencia original con desfases	Evita la pérdida de información por agregación preliminar	Requiere optimización no lineal no convexa en MIDAS
MF-VAR / MF-BVAR	VAR en Espacio de Estados con priores de contracción	Estimación estocástica de estados latentes vía MCMC	Modela interacciones endógenas y retroalimentación	Alto costo computacional; sensibilidad a la elección de priores

La frontera del aprendizaje automático (machine learning) y el procesamiento de datos no estructurados

La expansión de la capacidad de cómputo y la disponibilidad de bases de datos masivas (Big Data) impulsaron una transformación metodológica en el nowcasting macroeconómico mediante la incorporación de algoritmos de Aprendizaje Automático (Machine Learning o ML). Esta evolución responde a la necesidad de superar las dos limitaciones restrictivas de los modelos econométricos lineales: la incapacidad para gestionar sistemas donde la dimensión de los regresores excede el número de observaciones históricas ($p \gg n$) y la rigidez para capturar interacciones no lineales complejas entre las variables explicativas.

Para abordar el problema de la alta dimensionalidad sin recurrir a la extracción de factores abstractos, la literatura se inclinó hacia los métodos de regularización o regresiones penalizadas. Estas técnicas modifican la función de pérdida del modelo mediante la adición de una penalización a la magnitud de los coeficientes, forzando un encogimiento (shrinkage) que reduce la varianza del estimador y evita el sobreajuste (overfitting).

La función de optimización general para una regresión regularizada de pronóstico se define como (Camejo et al., 2019):

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{p} x_{tj}\, \beta_j \right)^2 + P(\beta; \lambda, \alpha)$$

Dentro de esta familia, Ridge Regression asigna una penalización L_2 equivalente a $P(\beta) = \lambda \sum_{j=1}^{p} \beta_j^2$, la cual comprime los coeficientes hacia cero sin anular ninguno por completo, manteniendo todos los regresores dentro de la ecuación. En contraste, la técnica LASSO (Least Absolute Shrinkage and Selection Operator), introducida por Tibshirani (1996), aplica una penalización L_1definida como $P(\beta) = \lambda \sum_{j=1}^{p} |\beta_j|$. Debido a la geometría de la norma L_1LASSO forzaba a que los coeficientes de las variables menos relevantes se anularan exactamente, realizando una selección automática de indicadores dentro del panel. Para corregir la inestabilidad de LASSO en presencia de regresores altamente correlacionados entre sí, Zou y Hastie (2005) formularon el algoritmo Elastic Net, el cual combina de forma ponderada ambas penalizaciones mediante la expresión $P(\beta) = \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2 \right]$, permitiendo seleccionar grupos de variables colineales de manera simultánea.

En el campo del modelado bayesiano de alta dimensión, la selección estocástica de regresores mediante métodos Spike-and-Slab (Scott y Varian, 2015) ofrece un marco probabilístico alternativo. Esta metodología impone una distribución a priori sobre los coeficientes que combina una masa puntual concentrada en cero (spike) y una distribución gaussiana difusa de alta varianza (slab). Mediante técnicas de muestreo de Gibbs, el algoritmo calcula la probabilidad posterior de inclusión de cada variable, seleccionando iterativamente las combinaciones de indicadores con mayor poder explicativo para predecir el PBI.

Para la captura de no linealidades y discontinuidades estructurales en el ciclo económico, los modelos basados en árboles de

decisión y ensambles han demostrado una notable flexibilidad. Algoritmos como Random Forest (RF) abordan la inestabilidad de los árboles individuales promediando los resultados de un conjunto masivo de árboles entrenados sobre submuestras aleatorias de datos y regresores (técnica de bagging), lo que reduce de forma drástica la varianza de la proyección. Por otro lado, los métodos de Gradient Boosting (como XGBoost y LightGBM) construyen árboles de decisión de forma secuencial, donde cada nuevo árbol se optimiza para ajustar los residuos o gradientes de error derivados de los árboles precedentes.

Complementariamente, la técnica de Regresión de Vectores de Soporte (SVR) minimiza el riesgo empírico construyendo un tubo insensible a margen ϵ alrededor de las observaciones, introduciendo variables de holgura (ξ_i, ξ_i^*) que permiten penalizar únicamente aquellas desviaciones que superan el umbral normativo establecido. Asimismo, las redes neuronales artificiales —especialmente las arquitecturas de Perceptrón Multicapa (MLP)— mapean interacciones complejas mediante capas ocultas compuestas por neuronas equipadas con funciones de activación no lineales.

Un avance sustancial en el nowcasting de última generación consiste en la integración de datos no estructurados de alta frecuencia para capturar señales económicas contemporáneas inaccesibles para la estadística tradicional. Entre estas fuentes destacan los índices de volumen de búsqueda extraídos de Google Trends, los cuales actúan como proxis en tiempo real de las decisiones de consumo, búsqueda de empleo e inversión privada. Del mismo modo, el procesamiento del lenguaje natural (NLP) aplicado al análisis de sentimientos de noticias de prensa económica permite cuantificar la

incertidumbre política y empresarial, influyendo sobre las predicciones de la actividad.

Estas fuentes no estructuradas se complementan con datos transaccionales de alta frecuencia, tales como registros informáticos de pagos digitales, uso de tarjetas bancarias, demanda eléctrica horaria y lecturas satelitales del tráfico de carga en nodos logísticos y puertos.

La evaluación rigurosa del desempeño predictivo fuera de muestra (out-of-sample) en modelos de nowcasting requiere esquemas de validación que respeten estrictamente el orden cronológico para evitar la filtración de información futura (data leakage). En lugar de la validación cruzada K -Fold convencional, la literatura utiliza esquemas de evaluación en tiempo real adaptados (Walk-Forward Cross-Validation o pseudo real-time nowcasting) mediante ventanas expandibles (expanding windows) o rodantes (rolling windows).

La cuantificación de la precisión del pronóstico se realiza mediante métricas estandarizadas en la econometría predictiva (Silva, 2013):

$$RMSE = \sqrt{\frac{1}{T_f}\sum_{t=1}^{T_f}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{T_f}\sum_{t=1}^{T_f}|\, y_t - \hat{y}_t|$$

$$SMAPE = \frac{100}{T_f}\sum_{t=1}^{T_f}\frac{2|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$$

$$MASE = \frac{\frac{1}{T_f}\sum_{t=1}^{T_f}|\,y_t - \hat{y}_t|}{\frac{1}{T-m}\sum_{t=m+1}^{T}|\,y_t - y_{t-m}|}$$

Para contrastar formalmente si las diferencias en la precisión predictiva entre dos modelos competidores son estadísticamente significativas, se aplica la prueba de Diebold-Mariano (1995) sobre la serie del diferencial de pérdidas $d_t = \ell(e_{1,t}) - \ell(e_{2,t})$. El estadístico Diebold-Mariano (DM), ajustado por autocorrelación muestral en horizontes $h > 1$, se distribuye asintóticamente bajo la hipótesis nula de igual precisión como una variable normal estándar $\mathcal{N}(0,1)$.

Análisis empírico comparativo y aplicaciones institucionales en bancos centrales

La implementación del nowcasting en la arquitectura de análisis macroeconómico de los bancos centrales ha dado lugar a configuraciones institucionales diferenciadas según la estructura productiva y la disponibilidad de información de cada país.

Un caso relevante de divergencia metodológica en una misma economía se observa al comparar la Reserva Federal de Nueva York y la de Atlanta. El NY Fed Staff Nowcast adopta un marco fuertemente conceptualizado en torno a un Modelo de Factores Dinámicos de carácter "Top-Down". Este modelo procesa decenas de indicadores macroeconómicos e infiere la trayectoria del PBI a través del comportamiento de un factor común latente que filtra el ruido aleatorio. Esta arquitectura garantiza estabilidad y suavizado en las estimaciones durante épocas de moderación económica, pero

presenta un tiempo de respuesta más lento ante choques atípicos de gran magnitud.

En contraposición, la herramienta GDPNow de la Reserva Federal de Atlanta, desarrollada por Higgins (2014), sigue un diseño "Bottom-Up" fundamentado en ecuaciones puente que proyectan de manera independiente las 13 subcomponentes del gasto del PBI real siguiendo la metodología de cuentas nacionales del Bureau of Economic Analysis (BEA). GDPNow incorpora de forma directa los datos crudos a medida que son publicados por las agencias oficiales sin aplicar filtros de suavizado ex-ante. Durante la crisis generada por la pandemia del COVID-19, la flexibilidad de GDPNow le permitió registrar de forma inmediata las contracciones y recuperaciones extremas de la actividad, mientras que los modelos factoriales tradicionales tendieron a amortiguar inicialmente el impacto debido a la inercia de sus factores latentes.

En América Latina, la elevada volatilidad macroeconómica, la exposición a choques de términos de intercambio y la presencia de sectores informales han motivado a las autoridades monetarias a desarrollar modelos adaptados a sus particularidades regionales.

En el Banco Central de Reserva del Perú (BCRP), la investigación empírica ha mostrado una evolución desde modelos de series de tiempo estructurales hasta la implementación de técnicas avanzadas de aprendizaje automático. Pérez (2020) desarrolló un modelo estructural de series de tiempo combinado con la técnica de Spike-and-Slab para el pronóstico del PBI mensual peruano, identificando que la producción de energía eléctrica, el volumen de importación de insumos industriales, el consumo interno de cemento y ciertos agregados del sistema financiero constituían los indicadores con

mayor probabilidad posterior de inclusión predictiva. Posteriormente, Tenorio y Pérez (2025) ampliaron esta frontera evaluando algoritmos de Machine Learning sobre una base de 53 variables estructuradas y 38 indicadores no estructurados (índices de sentimiento y Google Trends).

Sus resultados empíricos demostraron que los modelos de regresión regularizada (LASSO y Elastic Net) y los ensambles de Gradient Boosting alcanzaron una reducción en el error cuadrático medio ($RMSE$) de entre el 20% y el 25% en comparación con los modelos autorregresivos tradicionales AR(1) y las especificaciones DFM convencionales. De igual modo, Al comparar modelos Ridge, LASSO, Random Forest, XGBoost, SVR y redes neuronales (MLP) frente al benchmark DFM mediante esquemas de validación Walk-Forward, los autores confirmaron que las técnicas de Machine Learning logran una precisión predictiva superior al capturar patrones no lineales no contemplados por los modelos factoriales lineales. Adicionalmente, Winkelried y Urquizo (2026) incorporaron el análisis de las revisiones de datos (data vintages) para corregir sesgos predictivos sistemáticos introducidos por los procedimientos de empalme de la agencia estadística oficial.

En el Banco de México (Banxico), Gálvez-Soriano (2018, 2020) evaluó de forma comparativa cinco especificaciones de nowcasting para el PBI trimestral mexicano, incluyendo un Modelo de Factores Dinámicos (DFM), dos modelos de Ecuaciones Puente (BE) y dos aproximaciones basadas en Componentes Principales (PCA). Al realizar evaluaciones fuera de muestra en tiempo real y aplicar la prueba de Diebold-Mariano, el estudio demostró que el promedio de las Ecuaciones Puente superó estadísticamente la precisión del DFM, del

modelo PCA, de la mediana de los analistas del mercado financiero y de la estimación oportuna oficial publicada por el INEGI.

Experiencias similares se observan en el Banco Central del Uruguay (BCU), donde los modelos puente fundamentados en funciones de transferencia han demostrado un desempeño superior en horizontes de predicción inmediatos para estimar el valor agregado bruto sectorial (Brum y Rodríguez, 2016).

Por su parte, estudios elaborados por Espinosa (2020) determinaron que las curvas de ganancia acumulada y respuesta acumulada dan un enfoque práctico a los resultados del modelo elegido (XGBoost en este caso), el cual es importante en el ámbito del que proviene la base (sector financiero); por lo que se hacen conclusiones empíricas y de desempeño relativo en instituciones de EE.UU. y Perú (véase la tabla 5).

Tabla 5: Conclusiones empíricas y desempeño relativo en instituciones de EE.UU. y Perú

Institución / Autores	Variable Objetivo y Cobertura	Metodologías Evaluadas	Principales Regresores e Indicadores Clave	Conclusiones Empíricas y Desempeño Relativo
Reserva Federal de Nueva York	PBI Trimestral de EE. UU.	Modelo de Factores Dinámicos (DFM) Top-Down	Panel amplio de empleo, producción industrial, ventas y finanzas	Elevado suavizado y estabilidad en periodos regulares; respuesta gradual ante crisis
Reserva Federal de	PBI Trimestral de EE. UU.	Ecuaciones Puente Bottom-Up por	Datos crudos de comercio, construcción,	Alta reactividad en tiempo real; desempeño superior

Institución / Autores	Variable Objetivo y Cobertura	Metodologías Evaluadas	Principales Regresores e Indicadores Clave	Conclusiones Empíricas y Desempeño Relativo
Atlanta (Higgins)	(GDPNow)	components BEA	inventarios y consumo	durante choques como la pandemia
BCRP (Pérez-Forero, 2018)	PBI Mensual del Perú	Modelo Estructural de Series de Tiempo + Spike-and-Slab	Electricidad, importación de insumos, cemento y agregados crediticios	Selección probabilística automatizada de variables; generación de densidad de pronóstico
BCRP (Perez & Tenorio, 2024)	PBI Mensual del Perú	LASSO, Elastic Net, XGBoost vs. DFM y AR(1)	53 indicadores estructurados + 38 datos no estructurados (sentimiento/Google)	Reducción del $RMSE$ de entre 20% y 25% frente a DFM y AR(1); captura eficiente de no linealidades
BCRP (Flores et al., 2024)	PBI Total y No Primario del Perú	Ridge, LASSO, RF, XGBoost, SVR, MLP vs. DFM	Panel de más de 170 indicadores macroeconómicos y financieros	Los algoritmos ML superan al DFM benchmark mediante evaluación Walk-Forward
Banco de México (Gálvez-Soriano, 2020)	PBI Trimestral de México	DFM, Ecuaciones Puente (BE), PCA	Indicadores del IGAE, comercio exterior y demanda interna	El promedio de Ecuaciones Puente superó estadísticamente al DFM, encuestas y al INEGI oportuno
De Nederlandsche Bank (Pick et al., 2024)	PBI Trimestral de Países Bajos	DFM, MIDAS, LASSO, Elastic Net, Random Forest	Panel de indicadores macroeconómicos mensuales de la Eurozona	Random Forest ofreció el menor error en nowcasting; DFM destacó en backcasting

Implicancias operativas, transparencia e integración en la política monetaria

El despliegue de las herramientas de nowcasting en la rutina de análisis de los bancos centrales plantea dilemas operativos que van más allá del cómputo del error cuadrático medio. La función de un modelo de pronóstico inmediato en un instituto emisor consiste en fundamentar las deliberaciones del Comité de Política Monetaria mediante diagnósticos comprensibles y técnicamente sustentados.

En este contexto, la adopción creciente de modelos complejos de Machine Learning ha puesto en evidencia un trade-off analítico entre la capacidad predictiva fuera de muestra y la interpretabilidad de las proyecciones. Los Modelos de Factores Dinámicos y las Ecuaciones Puente tradicionales presentan una ventaja metodológica relevante: permiten ejecutar descomposiciones de "noticias" (news decomposition) (Brum y Rodríguez, 2016). Cuando se publica un nuevo dato económico, estos modelos no solo actualizan la estimación del PBI, sino que permiten aislar exactamente qué porcentaje de la revisión del pronóstico se debe a la desviación entre el dato observado y el valor que el modelo anticipaba (surprise component), así como la ponderación que el modelo otorga a esa variable específica.

Esta propiedad narrativa es vital para que los economistas del banco central puedan explicar la evolución de las proyecciones a las autoridades de la institución y a los agentes del mercado financiero.

Por el contrario, aunque las arquitecturas avanzadas de Machine Learning —tales como las Redes Neuronales o los ensambles de Gradient Boosting— suelen exhibir menor error predictivo durante periodos de volatilidad o en puntos de inflexión del ciclo, su naturaleza

no lineal dificulta la atribución aditiva transparente de las revisiones del pronóstico a indicadores individuales. Aunque herramientas de explicabilidad como los valores SHAP (Shapley Additive exPlanations) o la importancia de variables por permutación permiten aproximar la contribución marginal de los regresores, la reconstrucción exacta del flujo de noticias sigue siendo más compleja que en las estructuras econométricas lineales.

Frente a esta disyuntiva, la práctica contemporánea en la banca central ha convergido hacia el desarrollo de un marco integrado denominado enfoque de suite de modelos (suite-of-models approach). En este esquema de gestión de información, el banco central no depende de una única especificación econométrica, sino que opera un ecosistema diversificado de herramientas de forma paralela.

Los modelos DFM y de Ecuaciones Puente se utilizan como herramientas primarias de comunicación y descomposición contable del flujo de noticias macroeconómicas. Simultáneamente, los algoritmos de Aprendizaje Automático con regulación y no linealidad se emplean como sistemas de validación y detección de riesgos de cola, alertando tempranamente a los tomadores de decisión sobre posibles desviaciones o cambios de tendencia no capturados por los filtros lineales tradicionales (Brum y Rodríguez, 2016).

La metodología de pronóstico inmediato del Producto Bruto Interno se consolidó como una disciplina indispensable para la conducción de la política macroeconómica en entornos de información asíncrona. La evolución de este campo refleja una transición desde modelos autorregresivos univariados hacia sistemas multivariados de alta dimensión y algoritmos flexibles de aprendizaje Automático capaces de sintetizar señales dispersas en tiempo real.

Del análisis comparativo de la evidencia empírica internacional e institucional se desprenden tres conclusiones analíticas principales:

- En primer lugar, se demuestra que la idoneidad de un modelo de nowcasting depende de la fase del ciclo económico y de la naturaleza de las perturbaciones. Mientras que los Modelos de Factores Dinámicos ofrecen un desempeño eficiente y estable en periodos de volatilidad moderada gracias a su capacidad de reducción de ruido, las metodologías de Machine Learning (particularmente Gradient Boosting, LASSO y Elastic Net) superan a las especificaciones tradicionales durante fases de inflexión ciclo-económica y choques estructurales.
- En segundo lugar, la integración de datos no estructurados de alta frecuencia —tales como indicadores de sentimiento extraídos mediante NLP, métricas de búsqueda web y registros de pagos digitales— proporciona un valor predictivo incremental que complementa las series de tiempo de la estadística oficial. En tercer lugar, el marco operativo óptimo para un banco central exige la combinación de modelos a través de enfoques de suite, preservando la interpretabilidad necesaria para la comunicación de la política monetaria al tiempo que se maximiza la precisión predictiva fuera de muestra.

En adelante, la investigación sobre nowcasting macroeconómico se organiza en torno a tres fronteras emergentes:

1. Modelado en tiempo real de revisiones de datos (Data Vintages): El desarrollo de arquitecturas econométricas orientadas a capturar formalmente el proceso de revisión de las cuentas nacionales. Estos modelos no solo proyectan el PBI a partir de los indicadores disponibles, sino que corrigen los sesgos

sistemáticos derivados de las posteriores revisiones de las series por parte de las agencias de estadística oficial.

2. Modelos Fundacionales de Series de Tiempo (Time Series Foundation Models): La adaptación de redes profundas basadas en arquitecturas Transformer entrenadas sobre paneles masivos de datos macroeconómicos para ejecutar pronósticos inmediatos probabilísticos sin necesidad de reentrenamiento específico (zero-shot forecasting). Estos enfoques permiten evaluar la distribución completa del pronóstico y cuantificar riesgos de cola a través de métricas como el Continuous Ranked Probability Score (CRPS).

3. Sistemas Híbridos con Restricciones Estructurales: La combinación de algoritmos de Machine Learning no lineales con restricciones impuestas por identidades contables o relaciones macroeconómicas de equilibrio (provenientes de modelos DSGE de frecuencia mixta). Esta integración busca construir modelos de pronóstico que preserven la precisión no lineal del Aprendizaje Automático garantizando al mismo tiempo la coherencia teórica indispensable para la formulación de políticas públicas.

Capítulo 3: Modelos de proyección de recaudación tributaria: integración del PBI nominal, deflactores sectoriales y elasticidad fiscal

La precisión en la proyección de la recaudación tributaria constituye el pilar fundamental de la planificación macrofiscal, la formulación presupuestaria y la evaluación de la sostenibilidad de la deuda pública. En economías emergentes y en desarrollo, la estimación de los ingresos fiscales enfrenta retos metodológicos complejos derivados de la volatilidad macroeconómica, los choques en los precios de las materias primas y las transformaciones estructurales en los sectores productivos. Históricamente, las autoridades fiscales han empleado agregados macroeconómicos simplificados para prever la trayectoria de los ingresos del Estado. Sin embargo, el uso exclusivo del Producto Bruto Interno (PBI) real o agregado induce sesgos sistemáticos en las proyecciones al ignorar la evolución diferenciada de los precios sectoriales y las distintas sensibilidades impositivas según la rama de actividad económica.

El diseño moderno de modelos econométricos de proyección fiscal exige descomponer el crecimiento económico entre sus componentes reales y nominales, aplicando deflactores implícitos de precios a nivel macroeconómico y sectorial. Esta descomposición permite aislar la sensibilidad de los impuestos ante expansiones en el volumen de producción física frente a variaciones en el nivel general o específico de precios. Asimismo, la separación entre la elasticidad tributaria pura —que captura el comportamiento espontáneo del

sistema fiscal— y la flotabilidad o *boyanza* tributaria —que incorpora el efecto de las reformas legislativas discrecionales— resulta indispensable para la elaboración del Marco Macroeconómico Multianual y la gestión de las finanzas públicas (Conejo et al, 2011).

Fundamentos teóricos: elasticidad tributaria versus flotabilidad (Boyanza)

En la literatura econométrica y de finanzas públicas, la respuesta de los ingresos fiscales ante las fluctuaciones de la actividad económica se analiza a través de dos conceptos centrales: la flotabilidad (*tax buoyancy*) y la elasticidad tributaria (*tax elasticity*). La flotabilidad mide la variación porcentual total de la recaudación tributaria generada por un cambio porcentual en el PBI, reflejando el rendimiento observado sin ajustar las series históricas por los efectos de reformas tributarias discrecionales. Matemáticamente, la flotabilidad B se define mediante la expresión (Lahura y Castillo, 2016):

$$B = \frac{\Delta\%T}{\Delta\%Y} = \frac{\Delta T/T}{\Delta Y/Y}$$

Donde T representa la recaudación tributaria total o de un impuesto específico, y Y representa la base imponible macroeconómica, aproximada habitualmente por el PBI. Cuando $B > 1$, se sostiene que el sistema tributario es flotante, ya que la recaudación crece a una tasa proporcionalmente mayor que la producción nacional.

En contraste, la elasticidad tributaria (ε) cuantifica exclusivamente la respuesta automática del sistema fiscal ante variaciones en el producto, manteniendo constante la estructura legal, normativa e institucional de los impuestos. Este indicador aísla el impacto de las modificaciones en las alícuotas, los cambios en las

deducciones y exenciones, y las alteraciones en las bases imponibles (Vásquez, 2023). La estimación de la elasticidad resulta crítica para las proyecciones macrofiscales de mediano y largo plazo, dado que previene la extrapolación errónea de incrementos recaudatorios pasados originados en modificaciones legislativas de carácter transitorio. La relación de elasticidad se expresa formalmente como (Conejo et al., 2011):

$$\varepsilon = \frac{\Delta\%T}{\Delta\%Y}\Big|_{\text{política fiscal constante}}$$

Para depurar las series de recaudación y depurar los efectos discrecionales, la teoría econométrica aplica cuatro metodologías estandarizadas. El *método de ajuste proporcional* ajusta históricamente los ingresos descontando el rendimiento estimado de cada reforma legal en el año de su entrada en vigor y recalculando la serie de forma encadenada. El *enfoque de variables ficticias (dummy variables)* introduce variables dicotómicas de impulso o de cambio estructural en los modelos de regresión para capturar los quiebres normativos. El *método de estructura de tasa constante* aplica retroactivamente la estructura impositiva de un año base a la serie histórica de las bases imponibles. Por ende, el *método del índice Divisia* pondera el crecimiento de la recaudación mediante índices ponderados para separar el efecto producto de las alteraciones en la tasa efectiva promedio (Tapia, 2025).

En el contexto latinoamericano, los sistemas fiscales muestran una elasticidad de largo plazo superior a la unidad, impulsada principalmente por la dinámica del Impuesto a la Renta corporativo. Sin embargo, en el corto plazo, las elasticidades tienden a ser inelásticas o

a presentar comportamientos fuertemente asimétricos según la fase del ciclo económico (Sánchez, 2018).

La descomposición nominal vs. real: el impacto de la inflación y los deflactores

La mayoría de los impuestos que integran la estructura tributaria —incluyendo el Impuesto al Valor Agregado (IVA o IGV), los Impuestos Selectivos al Consumo (ISC), los aranceles y el Impuesto a la Renta de empresas y personas— se liquidan sobre bases imponibles expresadas en valores nominales o corrientes. Por consiguiente, proyectar los ingresos fiscales utilizando de manera exclusiva la tasa de crecimiento del PBI real introduce distorsiones sistemáticas, especialmente en contextos donde la tasa de inflación diverge de la meta de largo plazo (Instituto Nacional de Estadística e Informática, 1999).

El PBI nominal combina el volumen físico de bienes y servicios producidos (PBI real) con la variación en el nivel general de precios de la economía, representada por el Deflactor Implícito del PBI (DI). La relación matemática entre estos agregados se define formalmente como (INEI, 2025):

$$DI_t = \left(\frac{PBI_{nom,t}}{PBI_{real,t}} \right) \times 100$$

A diferencia del Índice de Precios al Consumidor (IPC), que mide la evolución de los precios de una canasta fija de bienes de consumo final, el Deflactor del PBI abarca las variaciones de precio de todos los bienes y servicios finales producidos internamente, incluyendo bienes de capital, insumos intermedios y exportaciones, al tiempo que excluye las importaciones.

Para estructurar modelos de proyección rigurosos, se especifican regresiones doble logarítmicas que separan explícitamente el PBI nominal en su componente de volumen y su componente de precio:

$$\ln(IT_t) = \alpha + \beta_1 \ln\left(PBI_{real,t}\right) + \beta_2 \ln(DI_t) + \mu_t$$

En esta especificación, IT_t representa la recaudación tributaria, β_1 expresa la elasticidad tributaria respecto al crecimiento real de la producción, y β_2 denota la elasticidad respecto al deflator de precios.

La evidencia empírica confirma que las elasticidades tributarias ante el producto real y ante la inflación no son simétricas ni unitarias. Los incrementos en la actividad económica real ejercen una influencia significativamente mayor sobre los ingresos fiscales que las variaciones puramente inflacionarias (Ervina et al., 2025). Por ejemplo, en el caso de Colombia, estimaciones econométricas señalan que la elasticidad de la recaudación tributaria respecto al PBI nominal alcanza el 1,10% (con cálculos alternativos en 1,17%).

Sin embargo, al descomponer el producto nominal, la elasticidad frente al PBI real asciende al 1,88%, mientras que la elasticidad frente a la inflación se sitúa en un 0,95%, demostrando que la recaudación responde de manera más enérgica al aumento en el volumen de actividad económica que a las presiones inflacionarias generales (véase la tabla 6).

Tabla 6: Ejemplos de elasticidad en la economía fiscal de Colombia y El Salvador

Economía / Indicador Fiscal	Elasticidad PBI Nominal Total	Elasticidad PBI Real (β1)	Elasticidad Deflactor Precios (β2)
Colombia (Recaudación Tributaria Total)	1,10% - 1,17%	1,88%	0,95%
El Salvador (Recaudación Tributaria Total)	—	1,75	1,23
El Salvador (Impuesto sobre la Renta - ISR)	—	2,00	1,72

Esta discrepancia entre β_1 y β_2 responde a mecanismos institucionales y rezagos de recaudación. Un aumento de la inflación eleva de forma inmediata las bases nominales del IVA; sin embargo, en presencia de plazos de pago y rezagos en la declaración impositiva, la inflación erosiona el valor real de las obligaciones tributarias (efecto Olivera-Tanzi). Adicionalmente, cuando las deducciones o los tramos impositivos se fijan en montos nominales sin indexación automática, la inflación desplaza artificialmente a los contribuyentes a tramos impositivos más altos (*bracket creep*), modificando la tasa efectiva de manera no lineal.

Deflactores sectoriales y desagregación por ramas de actividad

La utilización de un deflactor macroeconómico único asume que los precios de los distintos sectores de la economía evolucionan de manera uniforme. Dicha premisa resulta insostenible en economías

heterogéneas o primario-exportadoras, donde los términos de intercambio generan dinámicas de precios fuertemente dispares. Por ejemplo, el sector de minería e hidrocarburos puede experimentar un aumento drástico en su deflactor debido a choques de oferta internacionales, mientras que los sectores de servicios, construcción o agricultura registran variaciones moderadas en sus deflactores específicos (INEI, 2025).

Puesto que los sistemas tributarios no gravan con la misma intensidad a todas las ramas de la economía —la minería y las finanzas aportan una porción desproporcionada del Impuesto a la Renta de empresas, mientras que el comercio y la manufactura concentran la recaudación del IVA—, los modelos agregados incurren en desviaciones severas al proyectar ingresos. La metodología avanzada de proyección requiere especificar la función de recaudación en función del valor agregado nominal de cada sector, calculado mediante su respectivo deflactor sectorial ($DI_{i,t}$) Hendry y Ericsson (2003):

$$S_{i,t} = S_{i,real,t} \times \left(\frac{DI_{i,t}}{100}\right)$$

$$\ln(IT_t) = \beta_0 + \sum_{i=1}^{k} \beta_i \ln\left(S_{i,t}\right) + \mu_t$$

Estudios aplicados sobre la estructura fiscal de El Salvador evidencian la importancia de esta desagregación sectorial. La estimación econométrica para el total de ingresos tributarios arroja un vector de coeficientes altamente diferenciado por ramas de actividad:

$$\ln(IT_t) = -19{,}97 - 0{,}20 \ln\left(S_{1,t}\right) + 1{,}21 \ln\left(S_{2,t}\right) + 2{,}05 \ln\left(S_{3,t}\right)$$

El análisis de esta ecuación revela aspectos estructurales decisivos para la estimación de ingresos públicos. El Sector 1 (Agrícola)

presenta una elasticidad negativa y reducida (menos $-0{,}20$ para la recaudación total y menos $-0{,}54$ para el IVA), lo cual refleja los elevados niveles de informalidad rural, la presencia de exenciones sobre la canasta básica alimentaria y esquemas de tributación simplificada que desvinculan el valor producido de la obligación fiscal. Por el contrario, el Sector 2 (Industrial) muestra una elasticidad marcadamente positiva (1,21 para la recaudación total y 2,56 para el IVA), fundamentada en la alta formalización del sector manufacturero y su rol como agente de retención en la cadena de valor. Finalmente, el Sector 3 (Servicios) registra la mayor sensibilidad fiscal, con elasticidades de 2,05 respecto al ingreso total y 2,37 respecto al Impuesto sobre la Renta, atribuibles a la bancarización de las transacciones y la concentración de la fuerza laboral asalariada formal.

La incorporación de deflactores sectoriales permite proyectar el impacto de shocks externos selectivos. Un incremento en el deflactor de la minería expandirá la recaudación por el impuesto a las utilidades sin impactar mayormente el IVA al consumo interno, permitiendo desarmar los efectos puramente nominales e identificar la recaudación de carácter permanente (Balmore, 2023).

Arquitectura econométrica para la proyección macrofiscal

Para traducir estas relaciones en marcos operativos de proyección presupuestaria, se implementan modelos econométricos de series de tiempo capaces de capturar tanto la dinámica estacional de corto plazo como las relaciones de equilibrio estático de largo plazo.

Modelos de cointegración y corrección de errores (VECM y ARDL)

Debido a que las series de tiempo de recaudación tributaria y PBI nominal presentan habitualmente no estacionariedad —clasificándose como integradas de orden uno, $I(1)$—, el uso de regresiones lineales convencionales sobre variables en niveles puede resultar en relaciones espurias. Tras comprobar la presencia de raíces unitarias mediante la prueba de Dickey-Fuller aumentada (ADF), se verifica si existe al menos una relación de cointegración entre las variables (Mendoza, 2025).

Al confirmarse la cointegración, se especifica un modelo de corrección de errores (ECM) o una estructura de rezagos distribuidos autorregresivos (ARDL). La formulación general del modelo de corrección de error para la proyección de ingresos se expresa como:

$$\Delta \ln(IT_t) = \alpha + \sum_{j=1}^{p} \theta_j \, \Delta \ln\left(IT_{t-j}\right) + \sum_{k=0}^{q} \delta_k \, \Delta \ln(Y_{t-k}) + \lambda[\ln(IT_{t-1}) - \gamma \ln(Y_{t-1})] + \varepsilon_t$$

En esta arquitectura econométrica, el parámetro γ representa la elasticidad tributaria de largo plazo, el coeficiente δ_0 captura la elasticidad de corto plazo, y el parámetro λ mide la velocidad de ajuste del sistema tributario hacia su relación de equilibrio de largo plazo. Para garantizar la estabilidad del sistema, λ debe ser estadísticamente significativo y presentar signo negativo.

La aplicación de esta metodología en Bolivia demuestra la necesidad de diferenciar ambos horizontes temporales. La estimación de largo plazo arroja una elasticidad de la recaudación total respecto al PBI de 1,80, mientras que el IVA y el Impuesto a las Utilidades de las Empresas (IUE) presentan elasticidades de largo plazo de 2,29 y 1,93,

respectivamente. No obstante, en el corto plazo la sensibilidad aumenta sustancialmente: la elasticidad de corto plazo del IUE alcanza un valor de 9,9377 con un coeficiente de corrección de error de $-0{,}9997$, lo que indica un ajuste casi inmediato de la recaudación empresarial ante fluctuaciones en el producto nominal.

Asimetrías cíclicas y comportamiento no lineal

Las elasticidades tributarias no permanecen constantes a lo largo de las distintas fases del ciclo económico. Durante los periodos de expansión, las utilidades de las empresas se multiplican a tasas marginales superiores al crecimiento del PBI nominal, generando un comportamiento elástico de la recaudación. En contraste, en fases de desaceleración o recesión, la acumulación de pérdidas tributarias compensables reduce fuertemente la base imponible del Impuesto a la Renta.

En el caso del Perú, estudios econométricos sobre el Impuesto a la Renta de Personas Jurídicas (IR-PJ) confirman que en regímenes de expansión económica la elasticidad de corto plazo asciende a 2,48%, mientras que en fases recesivas la elasticidad disminuye drásticamente, modificando la trayectoria de la presión tributaria.

Modelación univariada y alta frecuencia (SARIMA)

Para el seguimiento de la ejecución presupuestaria intraanual y el pronóstico mensual de los ingresos, los modelos estructurales basados en el PBI se complementan con herramientas univariadas de series de tiempo de tipo SARIMA $(p, d, q)(P, D, Q)_s$. La recaudación mensual registra un acentuado componente estacional determinado por el calendario fiscal, observándose picos en los meses de

regularización anual del Impuesto a la Renta y caídas en los meses posteriores.

Evaluaciones empíricas a nivel de administraciones tributarias subnacionales y regionales —como las desarrolladas para los tributos internos administrados por la SUNAT en Puno— demuestran que especificaciones como SARIMA $(2,1,0)(1,1,1)_s$ satisfacen los supuestos de estacionariedad, normalidad e independencia de residuos, otorgando alta precisión para proyecciones de horizonte corto. (INEI, 2002)

Los enfoques econométricos para estimar la elasticidad tributaria a partir del PBI nominal, deflactores y desagregación sectorial han sido validados en diversos estudios empíricos en América Latina. La siguiente tabla sintetiza los principales hallazgos, metodologías y parámetros estimados en la región (véase la tabla 7):

Tabla 7: Evidencia empírica comparada en América Latina

País / Sistema Fiscal	Impuesto / Agregado	Elasticidad Corto Plazo	Elasticidad Largo Plazo	Deflactor o Variable de Precio	Metodología Econométrica
Colombia	Recaudación Total	—	1,10% - 1,17% (PBI Nom) 1,88% (PBI Real)	Elasticidad ante inflación: 0,95%	Modelo Estructural / Regresión Histórica
El Salvador	Recaudación Total	—	1,75 (PBI Real)	Deflactor Implícito PBI ($\beta = 1{,}23$)	Regresión Múltiple Logarítmica

País / Sistema Fiscal	Impuesto / Agregado	Elasticidad Corto Plazo	Elasticidad Largo Plazo	Deflactor o Variable de Precio	Metodología Econométrica
El Salvador	Impuesto Renta (ISR)	—	2,00 (PBI Real)	Deflactor Implícito PBI ($\beta = 1{,}72$)	Regresión Múltiple Logarítmica
El Salvador	IVA por Sectores	Sector 1: $-0{,}54$ Sector 2: $+2{,}56$ Sector 3: $-0{,}30$	—	Valor Agregado Sectorial Nominal	Regresión Sectorial Desagregada
Bolivia	Recaudación Total (R_t)	2,8852	1,8001	PBI Nominal General	Modelo Corrección de Error ($\lambda = -0{,}5296$)
Bolivia	IVA Interno	1,9793	2,2919	PBI Nominal General	Modelo Corrección de Error ($\lambda = -0{,}4923$)
Bolivia	Impuesto a Utilidades (IUE)	9,9377	1,9340	PBI Nominal General	Modelo Corrección de Error ($\lambda = -0{,}9997$)
Perú	Renta Sociedades (IR-PJ)	Regímenes de Expansión: 2,48%	—	PBI Nominal / Sectorial	Regresión Asimétrica no Lineal
Perú	Recaudación Interna Regional	—	—	Ajuste Estacional Interanual	Modelo SARIMA $(2,1,0)(1,1,1)_s$

En el Perú, el Marco Macroeconómico Multianual elaborado por el Ministerio de Economía y Finanzas (MEF) y los planes de gestión de

la SUNAT emplean estas estimaciones para proyectar una presión tributaria de equilibrio situada tradicionalmente entre 14,6% y 14,9% del PBI. En periodos recientes, divergencias entre el crecimiento del PBI nominal (+8,8%) y la recaudación efectiva han puesto de manifiesto la sensibilidad del sistema ante variaciones en los deflactores de materias primas y las dinámicas impositivas de los regímenes especiales.

Asimismo, las proyecciones impositivas peruanas deben ponderar la fragmentación de regímenes fiscales, como el Régimen Mype Tributario (RMT), con escalas progresivas del 10% (hasta 15 UIT) y 29,5% (sobre el exceso), el Régimen Especial del Impuesto a la Renta (RER) con alícuota del 1,5% sobre ingresos netos, y las modificaciones temporales en la tasa del IGV (como la reducción al 8% vigente para restaurantes y hoteles). La interacción entre estas estructuras normativas y las fluctuaciones en los deflactores sectoriales determina la trayectoria final de los ingresos fiscales netos.

La integración del PBI nominal, los deflactores sectoriales y la distinción entre elasticidad y flotabilidad dentro de los modelos de proyección de recaudación tributaria ofrece aportes metodológicos y prácticos para la formulación de la política fiscal.

En primer lugar, la utilización exclusiva del PBI real o de deflactores agregados para la estimación de ingresos tributarios introduce desviaciones sistemáticas en el presupuesto. Dado que las obligaciones fiscales se determinan sobre bases nominales, ignorar la heterogeneidad de los precios sectoriales conduce a errores en la previsión del espacio fiscal. Un shock positivo en el deflactor de las actividades extractivas genera un aumento transitorio en la recaudación por impuesto a la renta corporativa que no refleja una

mayor capacidad de gasto estructural ni una expansión permanente de la base imponible doméstica (INEI, 2025).

En segundo lugar, la constatación de que la elasticidad tributaria respecto al crecimiento económico real es superior a la elasticidad respecto a la inflación demuestra que la recaudación pública se consolida de manera más estable cuando el producto nominal es impulsado por la expansión de la producción física y la formalización laboral. Los incrementos de recaudación derivados de procesos inflacionarios no solo presentan menor intensidad proporcional, sino que se ven atenuados por los costos de recaudación y los rezagos en la liquidación de impuestos.

Finalmente, la adopción de arquitecturas econométricas multi-horizonte —que incorporen cointegración y modelos de corrección de error (ARDL/VECM) para las proyecciones de mediano plazo, modelos no lineales para capturar asimetrías cíclicas y especificaciones SARIMA para el control presupuestario mensual— constituye una práctica indispensable para dotar de sostenibilidad a las finanzas públicas. La correcta calibración de los deflactores sectoriales y las elasticidades asociadas fortalece la credibilidad de las reglas macrofiscales y asegura una gestión presupuestaria transparente ante choques económicos exógenos (Mendoza, 2025).

Capítulo 4: Modelos de elasticidad y comportamiento del consumo privado: microdatos de la Encuesta Nacional de Hogares (ENAHO) y variaciones del Índice de Precios al Consumidor (IPC)

El análisis del consumo privado y la estimación de la sensibilidad de los hogares ante variaciones en los precios relativos y el ingreso disponible constituyen instrumentos centrales para la formulación de políticas macroeconómicas y de bienestar. En economías emergentes caracterizadas por una elevada heterogeneidad en la distribución del ingreso y altos niveles de informalidad laboral, como la peruana, la modelación econométrica exige articular la microevidencia procedente de encuestas de hogares con los agregados macroeconómicos de inflación (INEI, 2017).

El uso de los microdatos de la Encuesta Nacional de Hogares (ENAHO), ejecutada de forma continua por el Instituto Nacional de Estadística e Informática (INEI), coordinado con las series regionales del Índice de Precios al Consumidor (IPC), permite estimar sistemas de demanda rigurosos, evaluar curvas de Engel no lineales y cuantificar los impactos distributivos desproporcionados derivados de choques inflacionarios (INEI, 2021).

Marco teórico y especificación de sistemas integrados de demanda

La teoría de la demanda microeconómica establece que las decisiones de consumo resultan de un proceso de optimización bajo

restricción presupuestaria. La estimación empírica de elasticidades precio propias, precio cruzadas e ingreso requiere la especificación de sistemas estructurales de demanda que respeten teóricamente las restricciones de aditividad, homogeneidad de grado cero en precios e ingreso, y simetría de las respuestas compensadas de Slutsky.

Evolución de los modelos econométricos de demanda

Las especificaciones uniecuacionales tradicionales, tales como los modelos lineales ($Q_t = \alpha_0 + \beta_1 P_t + u_t$) o dobles logarítmicos ($\ln Q_t = \alpha_0 + \beta_1 \ln P_t + u_t$), presentan limitaciones estructurales severas al asumir elasticidades constantes a lo largo de toda la curva de demanda o al violar las restricciones teóricas de agregación macroeconómica. Para superar estas deficiencias, la investigación empírica adopta sistemas de ecuaciones de demanda interdependientes.

El Sistema Casi Ideal de Demanda (AIDS, por sus siglas en inglés), desarrollado por Deaton y Muellbauer, representa las participaciones del gasto (w_i) en función de los precios relativos de la canasta y del gasto real total del hogar (INEI, 2025):

$$w_i = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{m}{a(P)}\right)$$

En esta especificación, m representa el gasto total nominal, p_j es el precio del bien j , y $a(P)$ representa un índice de precios translogarítmico, habitualmente aproximado en la práctica empírica por el índice de precios de Stone.

Para aproximar de manera más precisa el comportamiento del consumo en estratos de ingresos heterogéneos, Banks, Blundell y Lewbel introdujeron el Sistema Casi Ideal de Demanda Cuadrático (QUAIDS). Este modelo extiende el marco AIDS al incorporar el logaritmo del gasto real al cuadrado, capturando no linealidades esenciales en las curvas de Engel. De este modo, un bien puede comportarse como un bien de lujo a niveles bajos de ingreso y transformarse en un bien de primera necesidad a medida que el poder adquisitivo se incrementa:

$$w_i = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{m}{a(P)}\right) + \frac{\lambda_i}{b(P)}\left[\ln\left(\frac{m}{a(P)}\right)\right]^2$$

donde $b(P) = \prod_{k=1}^{N} p_k^{\beta_k}$ funciona como un deflator de escala del gasto. Alternativamente, el Sistema Lineal Extendido de Gasto (ELES), derivado de funciones de utilidad Stone-Geary, permite incorporar la decisión endógena de ahorro e identificar el consumo de subsistencia, aunque impone el supuesto restrictivo de aditividad estricta en las preferencias (véase la tabla 8).

Tabla 8: Evolución de los modelos econométricos de demanda

Modelo Econométrico	Especificación Funcional de Participación / Cantidad	Ventajas Metodológicas Principales	Restricciones Estructurales e Inconvenientes
MCO Log-Lineal	$\ln q_i = \alpha_0 + \beta \ln p_i + \gamma \ln m$	Estimación sencilla e interpretación directa de coeficientes.	Asume elasticidad constante y viola restricciones teóricas globales.
AIDS	$w_i = \alpha_i + \sum \gamma_{ij} \ln p_j + \beta_i \ln(m/P)$	Satisface propiedades teóricas de agregación e integración.	Impone curvas de Engel estrictamente lineales en el logaritmo del gasto.

Modelo Econométrico	Especificación Funcional de Participación / Cantidad	Ventajas Metodológicas Principales	Restricciones Estructurales e Inconvenientes
QUAIDS	$w_i = \alpha_i + \sum \gamma_{ij} \ln p_j + \beta_i \ln(m/P) + \frac{\lambda_i}{b(P)} [\ln(m /P)]^2$	Captura la no linealidad en curvas de Engel y cambios de estatus del bien.	Exige métodos de estimación no lineales e instrumentalización de variables.
Modelo de Deaton	$\ln v_{hc} = \alpha_0 + \beta \ln y_{hc} + \gamma z_{hc} + \ln p_c + u_{hc}$	Corrige simultáneamente el sesgo de calidad y la ausencia de precios observados.	Requiere el supuesto de variación de precios limitada al nivel de conglomerado.

Metodología econométrica para microdatos de hogares e indicadores de precios

La instrumentación de modelos de demanda a partir de microdatos de la Encuesta Nacional de Hogares (ENAHO) requiere abordar problemas metodológicos inherentes a las encuestas de corte transversal y panel. Entre estos destacan la ausencia de precios explícitos reportados por los consumidores, la presencia de ceros en las declaraciones de gasto (consumo censurado) y la disparidad del costo de vida entre ámbitos geográficos (INEI, 2021).

Construcción de precios relativos: valores unitarios y deflactación espacial

En la ENAHO, los datos primarios corresponden al gasto monetario total gastado en un producto (E_{ic}) y la cantidad física adquirida (q_{ic}) por el hogar i en el conglomerado c. El cociente directo entre ambas variables define el valor unitario ($v_{ic} = E_{ic}/q_{ic}$). Sin embargo, el valor unitario no es idéntico al precio de mercado, debido

a que incorpora decisiones endógenas sobre la calidad del bien. Los hogares de mayores ingresos suelen adquirir variedades de mayor precio unitario como reflejo de una calidad superior.

Para aislar la variación de precios puros, la metodología econométrica desarrollada por Deaton elimina las variaciones de calidad estimando ecuaciones de valores unitarios controladas por el ingreso del hogar (y_{ic}) y un conjunto de variables demográficas (z_{ic}), aislando la variación residual asociada al conglomerado espacial (p_c) (INEI, 2025):

$$\ln v_{ic} = \alpha_0 + \beta \ln y_{ic} + \gamma z_{ic} + \ln p_c + u_{ic}$$

Complementariamente, para articular los datos de la ENAHO con las variaciones temporales de las series del Índice de Precios al Consumidor (IPC) publicadas por el INEI, es indispensable aplicar deflactores espaciales de precios. Este procedimiento convierte los montos nominales de consumo en valores reales comparables entre dominios geográficos (Lima Metropolitana, Costa urbana/rural, Sierra urbana/rural y Selva urbana/rural), neutralizando las distorsiones que genera la disparidad en el costo de vida regional.

Corrección por sesgo de selección y consumo nulo

El reporte de cero gasto en determinados rubros de consumo durante el periodo de referencia de la encuesta constituye un problema sistemático. Este fenómeno responde a dos causas distintas: la decisión estructural del hogar de no consumir el producto por razones culturales o de preferencia (solución de esquina) o la falta de adquisición temporal dentro de la ventana de recolección de datos.

La estimación por MCO ignorando la censura introduce un sesgo de selección de muestra. Para corregirlo, la literatura aplica procedimientos en dos etapas basados en el modelo Probit (como las

metodologías de Heckman o Shonkwiler y Yen). En la primera etapa se determina la probabilidad condicional de que un hogar registre consumo positivo; posteriormente, se deriva la razón inversa de Mill o la función de distribución acumulada correspondiente para incorporarla como regresor de corrección en la estimación del sistema de demanda QUAIDS o AIDS.

Estimación empírica de elasticidades en la economía peruana

Las estimaciones empíricas derivadas de los microdatos de la ENAHO revelan una marcada heterogeneidad en la sensibilidad de la demanda según el tipo de producto, la posición en la distribución del ingreso y la zona geográfica.

Los productos alimenticios básicos de la canasta peruana muestran comportamientos inelásticos respecto al precio ($|E_{pp}| < 1$), lo que refleja su condición de bienes indispensables para la subsistencia familiar. Por el contrario, los bienes procesados, los servicios de entretenimiento, la educación privada y las bebidas alcohólicas exhiben demandas significativamente más elásticas ($|E_{pp}| > 1$).

Respecto a la respuesta frente al ingreso o gasto total (η_i), los alimentos primarios se clasifican como bienes normales de primera necesidad ($0 < \eta_i < 1$), mientras que las prendas de vestir, servicios diversos y bienes duraderos se ubican en la categoría de bienes superiores o de lujo ($\eta_i > 1$) (véase la tabla 9).

Tabla 9: Elasticidades precio e ingreso por agregados de consumo

Categoria / Producto de Consumo	Elasticidad Precio Propia (Epp)	Elasticidad Ingreso / Gasto (ηi)	Clasificación Microeconómica de Demanda
Arroz Corriente y Cereales Básicos	-0.20 a -0.35	0.20 a 0.45	Inelástico / Bien Normal Necesario.
Carne de Pollo / Menudencias	-0.65 a -0.90	0.60 a 0.85	Inelástico / Bien Normal Necesario.
Legumbres y Tubérculos (Lenteja, Ajo)	-0.30 a -0.50	0.10 a 0.35	Inelástico / Bien Normal Necesario.
Bebidas Alcohólicas (Cerveza)	-0.44 a -0.85	0.80 a 1.20	Inelástico a Elástico Moderado / Lujo Relativo.
Servicios de Recreación y Cultura	-1.10 a -1.50	1.20 a 1.60	Demanda Elástica / Bien Superior o de Lujo.
Servicios Públicos (Telefonía Móvil)	-0.35 a -0.65	0.65 a 0.95	Inelástico / Bien Normal Necesario.

Heterogeneidad por quintiles de ingreso y sustitución de bienes

La estructura presupuestaria de las familias peruanas varía drásticamente a lo largo de la distribución socioeconómica. Los hogares pertenecientes al Quintil I (el 20% más pobre de la población) destinan entre un 60% y un 70% de sus recursos totales exclusivamente a la adquisición de alimentos básicos. En contraste, en el Quintil V (el 20% más rico), dicha participación presupuestaria cae a rangos situados entre el 25% y el 30%.

Esta dispersión condiciona la magnitud de las elasticidades. Para los hogares en pobreza extrema, la elasticidad ingreso de los alimentos primarios es alta y cercana a la unidad, lo que implica que cualquier incremento marginal en los recursos se traduce casi en su totalidad en un aumento del consumo de calorías. Para los deciles superiores, la elasticidad ingreso de la nutrición básica cae por debajo de 0.20, derivándose las variaciones del gasto hacia atributos de mayor valor agregado o servicios fuera del hogar (Instituto Nacional de Estadística e Informática, 1999).

En materia de elasticidades cruzadas de la demanda (ε_{ij}), las estimaciones derivadas del sistema QUAIDS confirman patrones claros de sustitución dentro de la canasta alimentaria peruana. Ante incrementos imprevistos en los precios de productos como la carne de vacuno o el pescado, los hogares muestran una respuesta de sustitución positiva ($\varepsilon_{ij} > 0$) hacia opciones de menor costo por kilo, como la carne de pollo y las menudencias. Por el contrario, existen relaciones de complementariedad estricta ($\varepsilon_{ij} < 0$) entre insumos esenciales de preparación, como los aceites vegetales, sazonadores y cereales básicos.

Impacto distributivo de la inflación y evaluación de bienestar

El análisis de la interacción entre las elasticidades de demanda y las variaciones regionales del IPC permite medir las pérdidas de bienestar generadas por episodios inflacionarios sobre los presupuestos familiares.

Métricas de bienestar: variación compensada

La evaluación del impacto distributivo se realiza mediante el marco microeconómico de la Variación Compensada (CV). Esta métrica determina la compensación monetaria exacta requerida por un hogar para mantener su nivel de utilidad inicial tras experimentar una alteración en el vector de precios de la economía (INEI, 2025).

A partir de una aproximación de segundo orden de la función de gasto, la Variación Compensada expresada como proporción del gasto inicial del hogar (m_0) toma la siguiente especificación:

$$\frac{CV}{m_0} \approx \sum_{i=1}^{N} w_i \, \Delta \ln p_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \, \varepsilon_{ij}^{h} \Delta \ln p_i \, \Delta \ln p_j$$

donde w_i representa la participación inicial del bien i en el presupuesto familiar, $\Delta \ln p_i$ indica la variación porcentual del IPC específico del rubro i, y ε_{ij}^{h} denota la elasticidad precio cruzada compensada o Hicksiana derivada del sistema de demanda.

La aplicación de metodologías econométricas de Diferencias en Diferencias (*Diff-in-Diff*) combinadas con análisis de eventos sobre los microdatos de la ENAHO permite aislar el efecto causal de la inflación elevada sobre el consumo de los hogares peruanos. Los resultados empíricos demuestran que los choques inflacionarios operan con un carácter marcadamente regresivo.

Durante periodos de shock en los precios de los alimentos, los hogares situados en los quintiles inferiores experimentan reducciones severas y estadísticamente significativas en su gasto real per cápita, alcanzando caídas de hasta un **8,6%**. En contraste, los hogares con

mayores recursos logran amortiguar el impacto sobre su consumo sin registrar variaciones sustanciales en su nivel de vida.

La modelación de las decisiones de gasto bajo la Hipótesis del Ingreso Permanente (PIH) aclara la causa de esta divergencia. Los hogares de mayores ingresos presentan una propensión marginal a consumir (PMC) respecto al ingreso permanente sustancialmente superior a la correspondiente al ingreso transitorio. Esto confirma su capacidad para suavizar el consumo mediante la liquidación de ahorro acumulado, la reasignación de activos financieros o el acceso a mecanismos de crédito formal.

Por su parte, los hogares vulnerables enfrentan restricciones severas de liquidez. Ante la ausencia de ahorro o acceso al sistema financiero formal, su propensión a consumir del ingreso transitorio iguala a la del ingreso permanente, lo que los fuerza a ajustar el choque inflacionario de forma inmediata recortando el volumen físico de alimentos comprados (véase la tabla 10).

Tabla 10: Evidencia empírica de choques inflacionarios en el Perú

Quintil de Ingreso Socioeconómico	Proporción de Gasto en Alimentos (walim)	Variación del Gasto Real ante Inflación Alta	Mecanismo de Ajuste Predominante frente al IPC	Capacidad de Suavizamiento del Consumo
Quintil I (Muy Bajo)	> 60% del presupuesto	−8.6% a −6.5%	Contracción directa en cantidad y calidad alimentaria.	Nula (restricción de liquidez y sin crédito).
Quintil II - III (Medio-Bajo)	45% a 55% del presupuesto	−4.5% a −3.0%	Sustitución de bienes por variedades de menor precio.	Limitada (ahorro e inserción

Quintil de Ingreso Socioeconómico	Proporción de Gasto en Alimentos (walim)	Variación del Gasto Real ante Inflación Alta	Mecanismo de Ajuste Predominante frente al IPC	Capacidad de Suavizamiento del Consumo
				laboral secundaria).
Quintil IV - V (Alto)	$< 30\%$ del presupuesto	-1.0% a 0.0%	Reasignación de partidas de gasto discrecional y ahorro.	Elevada (Ajuste por activos e ingreso permanente).

Dinámica de ajuste intertemporal: oferta laboral y flexibilidad del ingreso

Las decisiones de consumo e ingreso están interconectadas. Ante la pérdida de poder adquisitivo provocada por variaciones no anticipadas en el IPC, los hogares peruanos no solo ajustan sus decisiones de demanda de bienes, sino que modifican su oferta de horas de trabajo para amortiguar el impacto sobre el bienestar familiar.

La elasticidad de la oferta laboral de Frisch mide la variabilidad de las horas trabajadas frente a cambios en el salario real por hora, manteniendo constante la utilidad marginal de la riqueza (λ):

$$\ln h_i = \psi \ln w_i + \psi \ln \bar{\lambda} - \psi \ln B$$

Estimaciones desarrolladas a partir de paneles rotativos y encuestas de empleo alineadas con la ENAHO ubican la elasticidad de Frisch promedio para la economía peruana en **0.38**, empleando estimadores de efectos fijos con variables instrumentales y corrección por sesgo de selección. Sin embargo, la estructura del mercado laboral peruano introduce asimetrías profundas en esta respuesta:

1. Los trabajadores masculinos en el sector formal presentan una oferta laboral inelástica (con valores situados entre -0.08 y 0.18), con reducido margen para incrementar horas de trabajo en su ocupación principal ante una aceleración del IPC.

2. La oferta laboral de las mujeres e independientes en el sector informal exhibe una elasticidad significativamente más elevada y flexible (alcanzando valores entre 0.40 y 2.79).

Dado que la tasa de informalidad laboral a nivel nacional supera el 75%, el ajuste del ingreso familiar frente a la inflación se genera mayoritariamente mediante la inserción de miembros secundarios del hogar en actividades informales no asalariadas o a través del incremento de horas en empleos secundarios. Si bien este mecanismo de generación de ingreso transitorio mitiga temporalmente la caída del consumo, perpetúa la precarización del empleo y expone a los hogares a choques de productividad de largo plazo.

Implicancias de política económica y protección social

La integración de la modelación de elasticidades de demanda con la microevidencia de la ENAHO proporciona bases empíricas para la optimización de instrumentos de política fiscal, monetaria y de asistencia social.

Eficiencia de transferencias monetarias frente a subsidios indirectos

La elevada pérdida de bienestar en los quintiles de menores ingresos ante aceleraciones en el IPC de alimentos respalda la prioridad de implementar transferencias monetarias directas y bien focalizadas, en lugar de recurrir a exenciones tributarias o subsidios de

precios generalizados (tales como exoneraciones del Impuesto General a las Ventas - IGV).

Dado que la elasticidad ingreso de los productos primarios en el Quintil I es alta, el otorgamiento de un subsidio directo restaura la capacidad de consumo real sin distorsionar la señal de precios relativos en el mercado ni generar pérdidas ineficientes de recaudación en la caja fiscal.

Diseños de Impuestos Selectivos al Consumo (ISC)

La estimación precisa de elasticidades-precio mediante el sistema QUAIDS resulta fundamental para calibrar las tasas del Impuesto Selectivo al Consumo (ISC) aplicadas a bienes con externalidades negativas, como las bebidas alcohólicas, el tabaco o los combustibles.

Al constatarse empíricamente que la demanda de cerveza presenta una elasticidad precio inelástica en el corto plazo (situada entre -0.44 y -0.85), un incremento en la tasa impositiva genera incrementos netos en la recaudación fiscal sin inducir caídas desproporcionadas en las ventas totales, cumpliendo su doble propósito fiscal y desincentivador de externalidades.

Transmisión monetaria y anclaje de expectativas

Para la conducción de la política monetaria por parte del Banco Central de Reserva del Perú (BCRP), comprender las heterogeneidades en la elasticidad del consumo resulta indispensable para proyectar la intensidad del traspaso inflacionario (*pass-through*) y evaluar los mecanismos de transmisión de la tasa de interés de referencia.

Dado que los estratos de menores ingresos carecen de instrumentos de ahorro para suavizar su consumo, los choques de precios ejercen una presión contractiva sobre la demanda agregada. Garantizar el anclaje de las expectativas inflacionarias dentro del rango meta (1% – 3%) constituye la política de protección social más eficaz para preservar el consumo real y evitar deterioros regresivos en la distribución del bienestar nacional.

La modelación econométrica del consumo privado en el Perú demuestra la necesidad de incorporar la heterogeneidad de los hogares para comprender la dinámica de la demanda ante variaciones en los precios e ingresos. La estimación de sistemas de demanda flexibles como el QUAIDS a partir de los microdatos de la ENAHO confirma que los bienes alimenticios básicos poseen elasticidades precio inelásticas y altas elasticidades ingreso en los deciles inferiores, lo que convierte a la inflación en un fenómeno regresivo que afecta en mayor medida a las familias vulnerables.

La ausencia de mecanismos de suavizamiento del consumo en estos estratos, provocada por restricciones de liquidez y una alta informalidad laboral, acentúa la contracción del gasto real frente a incrementos en el IPC. Por consiguiente, la formulación de políticas públicas debe basarse en la calibración empírica de estas elasticidades, priorizando esquemas de protección social focalizados y manteniendo la estabilidad macroeconómica como condición indispensable para preservar el bienestar de los hogares.

Capítulo 5: Modelos de riesgo crediticio y estimación de la probabilidad de incumplimiento en el Banco de Crédito del Perú

La gestión del riesgo de crédito en el sistema financiero latinoamericano ha experimentado una profunda transformación estructural, impulsada por la convergencia de modelos econométricos avanzados, arquitecturas de aprendizaje automático y un riguroso marco regulatorio dual. En el ámbito peruano, el Banco de Crédito del Perú (BCP) destaca como la entidad líder del sistema financiero, desplegando sofisticadas metodologías para la estimación de la Probabilidad de Incumplimiento (*Probability of Default* o PD), componente angular en la determinación de las pérdidas esperadas y en la asignación eficiente de capital.

El modelamiento de la PD en el BCP abarca desde la evaluación tradicional de solvencia hasta la explotación masiva de señales transaccionales de alta frecuencia, integrando variables del historial crediticio interno, dinámicas de saldos de cuenta, patrones temporales de morosidad y niveles de apalancamiento multifinanciero. La coexistencia de las exigencias prudenciales de la Superintendencia de Banca, Seguros y AFP (SBS) y las Normas Internacionales de Información Financiera (NIIF 9 / IFRS 9) a nivel del grupo Credicorp exige una calibración técnica precisa entre las métricas de provisión regulatoria y las pérdidas esperadas contables (Breed et al., 2020).

Arquitectura de modelamiento crediticio y factores determinantes de la PD

La estimación de la probabilidad de incumplimiento se fundamenta en la cuantificación del riesgo de que un deudor no cumpla con sus obligaciones contractuales en un horizonte temporal determinado. En las metodologías econométricas utilizadas por el BCP y supervisadas por el Banco Central de Reserva del Perú (BCRP), la regresión logística (*Logit*) se ha mantenido como el estándar analítico para la estructuración de tarjetas de puntuación y curvas de calibración. La especificación formal de la probabilidad condicional de incumplimiento se define mediante la función logística:

$$P(Y_i = 1 \mid X_i) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij}\right)}}$$

En esta formulación, $Y_i = 1$ denota la ocurrencia del evento de incumplimiento para el cliente i, X_{ij} representa la j-ésima variable predictiva del perfil financiero del cliente, y β_j corresponde a los coeficientes estimados por el método de máxima verosimilitud. Alternativamente, en la formulación de modelos *logit* sobre la razón de posibilidades (*odds ratio*), la estructura predictiva del banco evalúa la relación funcional entre los factores internos y externos del cliente:

$$\ln\left(\frac{PD_i}{1 - PD_i}\right) = \beta_0 + \beta_1 \text{HIST}_{12M,i} + \beta_2 \text{BAL}_{vol,i} + \beta_3 \text{MORA}_{dias,i} + \beta_4 \text{APAL}_{ratio,i} + \varepsilon_i$$

En la práctica contemporánea de la banca múltiple, el paradigma econométrico clásico es complementado por algoritmos no lineales de gradiente impulsado (*Gradient Boosted Decision Trees*), con especial énfasis en arquitecturas LightGBM. Para mitigar la opacidad analítica de estas redes de decisión y responder a los requerimientos de

auditabilidad de la Unidad de Riesgos de Crédito, la interpretabilidad del modelo se realiza mediante la descomposición de valores SHAP (*SHapley Additive exPlanations*), permitiendo medir la contribución marginal exacta de cada predictor en la PD predicha:

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j \, x'_j$$

Donde ϕ_0 representa el valor base esperado del modelo y ϕ_j denota la atribución de la variable j-ésima en la probabilidad predicha para una exposición específica.

Dentro de esta arquitectura multivariada, el historial crediticio interno del cliente constituye la variable de mayor peso predictivo en los modelos de *Behavior Scoring*. El análisis continuo del comportamiento histórico en las operaciones con el banco permite registrar la severidad y frecuencia del uso de líneas, distinguiendo entre clientes que cancelan la totalidad de su saldo mensual y aquellos que realizan únicamente el pago mínimo, comportamiento que incrementa la PD condicional. La evidencia empírica confirma que un deterioro en el patrón de pagos internos anticipa la transición hacia el incumplimiento de forma sensiblemente más ágil que las marcas informadas por centrales de riesgo externas.

El monitoreo de los saldos de cuentas y la volatilidad de la liquidez operan como un mecanismo de alerta temprana antes de que se formalice un impago contractual. A través del rastreo de los saldos medios en cuentas corrientes y de ahorros, el modelo evalúa la variabilidad del flujo de depósitos y el ratio entre el saldo disponible y la cuota mensual del crédito. Una caída abrupta o una volatilidad inusual en los depósitos refleja tensiones inmediatas en la caja del

deudor, elevando la PD estimada incluso cuando el cliente se encuentra al día en sus obligaciones.

Respecto a los **retrasos en pagos y la morosidad acumulada**, los modelos cuantifican no solo los días de atraso corrientes, sino también la métrica de recaídas en ventanas de 12 meses. La varianza temporal en las fechas de pago respecto al vencimiento oficial constituye un síntoma de iliquidez estructural, incrementando exponencialmente la PD en cuanto los atrasos superan los umbrales críticos de 8, 30 o 60 días.

Por último, el **nivel de apalancamiento y sobreendeudamiento** mide la exposición global del cliente frente al sistema financiero consolidado mediante el Reporte Crediticio Consolidado (RCC) de la SBS. En los segmentos de crédito revolvente y consumo, la tasa de utilización de la línea concedida exhibe una relación positiva monótona con la probabilidad de default. Los deudores situados en los deciles superiores de uso de línea, que presentan además endeudamiento en múltiples instituciones, muestran una probabilidad de incumplimiento sustancialmente más elevada que aquellos con líneas subutilizadas y concentradas en una sola entidad.

Segmentación de portafolio, variables transaccionales y scoring dinámico

El marco analítico de riesgo de crédito en el BCP opera mediante una clara segregación funcional entre el *Application Scoring*, enfocado en la admisión y originación mediante capacidad de pago demostrada o inferida, y el *Behavior Scoring*, orientado al monitoreo y recalibración periódica del portafolio activo (véase la tabla 11).

Tabla 11: Marco analítico de riesgo de crédito en el BCP

Categoría de Predictor	Variable Específica	Tipo de Scoring	Impacto Relativo en PD (SHAP / Logit)	Mecanismo de Transmisión al Riesgo
Historial Regulatorio	MÁX_DIAS_ATRASO_12M	Behavior Scoring	Muy Alto (Positivo)	Refleja la disciplina de pago y la recurrencia de shocks de liquidez previos.
Comportamiento Interno	RATIO_PAGO_MINIMO	Behavior Scoring	Alto (Positivo)	La dependencia del pago mínimo evidencia restricciones en el flujo de caja libre.
Apalancamiento Sistémico	NUM_ENTIDADES_RCD	Admisión / Behavior	Alto (Positivo)	La dispersión del endeudamiento eleva la vulnerabilidad ante shocks sistémicos.
Liquidez Transaccional	SALDO_MEDIO_MES / CUOTA	Behavior Scoring	Alto (Negativo)	La cobertura holgada de cuotas mediante saldos a la vista reduce la PD.
Uso de Crédito Revolvente	RATIO_USO_LINEA	Behavior Scoring	Moderado-Alto (Positivo)	La saturación de líneas revolventes anticipa estrés financiero latente.

Categoría de Predictor	Variable Específica	Tipo de Scoring	Impacto Relativo en PD (SHAP / Logit)	Mecanismo de Transmisión al Riesgo
Inteligencia Digital (Yape)	VELOCIDAD_TRANSACCIONAL	Micro-Scoring	Moderado (Negativo)	El flujo recurrente de cobros digitales valida ingresos en deudores no bancarizados.

La masificación de la billetera digital Yape dentro del ecosistema del BCP ha transformado la evaluación del riesgo crediticio en los segmentos micro-retail e informales. Mediante arquitecturas analíticas alimentadas por datos transaccionales de alta frecuencia, la entidad infiere la capacidad real de generación de ingresos sin depender de documentación contable tradicional. El volumen, la frecuencia y la diversidad de contrapartes en las operaciones procesadas por Yape se integran en modelos de micro-scoring que ajustan dinámicamente los límites de crédito y la PD asignada.

Esta recalibración de modelos y la optimización de las pautas de admisión han permitido una mejora en la calidad de la cartera de consumo y microempresa. Como resultado, la cartera de alto riesgo del BCP descendió desde un pico del 6.4% a inicios de 2024 hasta un 4.9% a mediados de 2025, favoreciendo la normalización de la prima por riesgo del banco hacia niveles del 1.7%.

Marco regulatorio de la SBS y comparativa con el estándar internacional IFRS 9

El BCP gestiona su riesgo de crédito bajo un esquema de doble cumplimiento normativo. A nivel local, cumple estrictamente las normas prudenciales dictadas por la SBS mediante el Reglamento para la Evaluación y Clasificación del Deudor y la Exigencia de Provisiones (Resolución SBS N° 11356-2008 y sus modificatorias). Simultáneamente, para la consolidación de sus estados financieros dentro del grupo Credicorp, aplica los estándares internacionales de la NIIF 9 / IFRS 9, basados en modelos de Pérdida Esperada (*Expected Credit Loss* o ECL) prospectivos.

Categorización regulatoria de deudores según la SBS

La normativa prudencial peruana exige agrupar a la totalidad de los deudores en cinco categorías de riesgo. Para la cartera minorista (créditos de consumo revolvente, consumo no revolvente e hipotecario para vivienda), la clasificación se realiza objetivamente sobre la base de los días calendario de atraso registrados. La categoría Normal comprende a deudores con atrasos de 0 a 8 días; la categoría Con Problemas Potenciales (CPP) abarca a aquellos con atrasos de 9 a 30 días; la categoría Deficiente agrupa a deudores con mora de 31 a 60 días; la categoría Dudoso incluye a quienes registran atrasos de 61 a 120 días; y la categoría Pérdida se asigna a deudores con impagos superiores a 120 días.

Para la cartera no minorista (créditos corporativos, a grandes empresas y a medianas empresas), la clasificación responde fundamentalmente al análisis de la capacidad de pago y la solvencia del flujo de caja del deudor, utilizándose los días de mora como un

criterio complementario. Un deudor no minorista es clasificado en la categoría Normal si presenta flujos de caja holgados y atrasos de 0 a 15 días. Se asigna la categoría CPP ante inconvenientes puntuales en la capacidad de pago o retrasos de 16 a 60 días. La clasificación Deficiente aplica a empresas cuyos flujos operativos no alcanzan a cubrir el servicio de la deuda o registran morosidad de 61 a 120 días. La categoría Dudoso corresponde a deudores con pérdidas operativas severas o atrasos de 121 a 365 días. Finalmente, la categoría Pérdida se otorga a empresas en estado de insolvencia manifiesta, con patrimonio neto negativo o con impagos que exceden los 365 días.

Adicionalmente, la SBS impone la aplicación del Criterio de Alineamiento Consolidado. En virtud de este mecanismo, si un deudor mantiene obligaciones con múltiples entidades del sistema financiero peruano, el BCP debe ajustar la clasificación del cliente a la categoría de mayor riesgo asignada por cualquier otra institución, siempre que se alcancen las proporciones de deuda reguladas. En el segmento de deudores minoristas, el alineamiento obligatorio se activa de forma automática cuando el deudor es clasificado como Dudoso o Pérdida en otra entidad (véase la tabla 12).

Tabla 12: Comparación entre categorizaciones regulatorias de deudores BCP y IFRS 9

Dimensión Analítica	Marco Regulatorio SBS (Resolución N° 11356-2008)	Estándar Internacional IFRS 9 (NIIF 9)
Propósito Principal	Cobertura prudencial de solvencia y preservación de capital.	Valuación razonable y reconocimiento oportuno del deterioro.

Dimensión Analítica	Marco Regulatorio SBS (Resolución N° 11356-2008)	Estándar Internacional IFRS 9 (NIIF 9)
Segmentación por Estado	5 Categorías: Normal, CPP, Deficiente, Dudoso, Pérdida.	3 Estadios: Stage 1 (*Performing*), Stage 2 (*Underperforming*), Stage 3 (*Default*).
Horizonte de la PD	PD implícita regulada mediante tasas porcentuales fijas de provisión.	Stage 1: PD a 12 meses. Stage 2 y 3: PD *Lifetime* (Vida Residual).
Inclusión Macroeconómica	Regla de provisión procíclica activada según el crecimiento del PBI.	Escenarios prospectivos múltiples ponderados (*Forward-Looking*).
Tratamiento de Garantías	Reducción directa de la tasa de provisión específica (Tablas 1, 2 y 3).	Modificación dinámica de la Severidad de Pérdida (LGD = 1 – Tasa de Recuperación).

Para atenuar las exigencias de provisiones específicas sobre deudores deteriorados, el marco regulatorio de la SBS reconoce el efecto mitigante de las garantías preferidas debidamente constituidas e inscritas (véase la tabla 13).

Tabla 13: Mitigación por garantías preferidas según la SBS

Categoría de Riesgo del Deudor	Tabla 1: Sin Garantía Preferida	Tabla 2: Con Garantía Preferida (Ej. Hipoteca)	Tabla 3: Con Garantía Preferida de Muy Rápida Realización
Normal	1.00%	1.00%	1.00%
Con Problemas Potenciales (CPP)	5.00%	2.50%	1.25%
Deficiente	25.00%	12.50%	15.00%

Categoría de Riesgo del Deudor	Tabla 1: Sin Garantía Preferida	Tabla 2: Con Garantía Preferida (Ej. Hipoteca)	Tabla 3: Con Garantía Preferida de Muy Rápida Realización
Dudoso	60.00%	30.00%	15.00%
Pérdida	100.00%	60.00%	30.00%

Determinación de pérdidas esperadas bajo IFRS 9

El cálculo contable de provisiones bajo NIIF 9 en Credicorp y BCP se estructura mediante la estimación probabilística de la Pérdida Esperada (ECL), proyectada a lo largo del ciclo de vida del crédito de Breed (2020):

$$ECL = \sum_{t=1}^{T} \frac{PD_t \times LGD_t \times EAD_t}{(1+r)^t}$$

Donde EAD_t denota la exposición al momento del incumplimiento (*Exposure at Default*), LGD_t representa la Severidad de la Pérdida (*Loss Given Default*) y r corresponde a la tasa de interés efectiva.

El elemento crítico bajo NIIF 9 es la identificación del Incremento Significativo en el Riesgo de Crédito (*Significant Increase in Credit Risk* o SICR). Cuando una exposición experimenta un SICR desde su originación, la operación transita automáticamente de **Stage 1** (donde se provisiona la Pérdida Esperada a 12 meses) hacia **Stage 2** (donde se exige la provisión de la Pérdida Esperada *Lifetime* o por la vida residual total del crédito). El BCP define el SICR combinando umbrales cuantitativos de incremento relativo en la PD marginal, mora mayor a 30 días o degradación en la calificación crediticia interna. Por su parte,

el **Stage 3** agrupa a todas las exposiciones en default de pagos (morosidad superior a 90 días o evidencia objetiva de deterioro financiero).

El desempeño del portafolio crediticio del Banco de Crédito del Perú evidencia la solidez de sus modelos de originación y la efectividad de sus mecanismos de cobertura ante ciclos económicos adversos (véase la tabla 14).

Tabla 14: Evolución del perfil de riesgo, indicadores de salud financiera y sensibilidad macroeconómica

Métrica Financiera / Indicador de Riesgo	Valor Reportado / Desempeño Reciente	Significación Operativa y Financiera
Rentabilidad sobre Activos (ROA)	2.8%	Nivel superior al promedio del sistema financiero (2.0%), apoyado en márgenes y eficiencia operativa.
Margen Financiero Bruto	75.7%	Expansión sólida frente al 73,0% previo, reflejando optimización en la estructura de fondeo.
Ratio de Cartera de Alto Riesgo	4.9% - 5.7%	Contracción sostenida desde techos del 6,5%, confirmando la efectividad de la calibración de modelos.
Variación Anual de Castigos	-31.7%	Caída pronunciada en la ejecución de castigos definitivos por incobrabilidad.
Ratio de Castigos / Créditos Directos	1.53%	Reducción del perfil de pérdida directa sobre la cartera total colocada.
Ratio de Liquidez (Moneda Nacional)	27.89%	Holgada posición de liquidez para afrontar requerimientos operativos y regulatorios.

Métrica Financiera / Indicador de Riesgo	Valor Reportado / Desempeño Reciente	Significación Operativa y Financiera
Prima por Riesgo	1.7%	Normalización del costo de riesgo respaldada por la mejora en la originación minorista.

La estimación de la PD dentro de los modelos condicionales del banco incorpora la sensibilidad del portafolio ante fluctuaciones del entorno macroeconómico. La evolución del Producto Bruto Interno (PBI) mantiene una relación inversa con la probabilidad de default; periodos de desaceleración productiva reducen los flujos operativos de las empresas medianas y pymes, elevando las tasas de transición hacia categorías de riesgo y acelerando migraciones hacia Stage 2 en la cartera comercial.

El tipo de cambio Sol/Dólar representa otra variable clave en el monitoreo del riesgo cambiario crediticio. Aquellos deudores que perciben sus ingresos en moneda nacional pero mantienen créditos en dólares experimentan un incremento directo en su PD ante episodios de depreciación del sol, debido al aumento en la carga financiera relativa sobre sus flujos semanales o mensuales.

Asimismo, la trayectoria de la tasa de referencia del BCRP influye sobre el costo de financiamiento de las líneas de consumo y créditos revolventes. Reducciones graduales en la tasa de política monetaria disminuyen la presión sobre el servicio de la deuda, atenuando la probabilidad de impago en las carteras de menor capacidad de soporte.

La arquitectura de gestión del riesgo de crédito en el Banco de Crédito del Perú representa una plataforma analítica integral que

equilibra el rigor econométrico con las capacidades del aprendizaje automático moderno. La coexistencia de modelos supervisados de regresión logística para fines regulatorios y modelos basados en algoritmos LightGBM explicados mediante SHAP permite al banco optimizar la precisión predictiva de la PD sin comprometer la transparencia exigida por la SBS.

El aprovechamiento estratégico de variables internas de alta frecuencia —tales como la volatilidad de depósitos, los patrones de pago mínimo y la inteligencia de flujos generada por la plataforma Yape— ha dotado al BCP de una capacidad adelantada para detectar señales de sobreendeudamiento e iliquidez antes de que estas se traduzcan en morosidad contractual. Esta ventaja analítica ha sido determinante en la reducción sostenida del ratio de cartera de alto riesgo y en la normalización de la prima por riesgo del negocio.

Por último, la sincronización entre las exigencias regulatorias locales basadas en categorías fijas de deudor y el estándar contable prospectivo de la NIIF 9 fortalece el perfil de solvencia de la institución. La estructura de cobertura, respaldada por un sólido margen financiero bruto, posiciones holgadas de liquidez y una tendencia decreciente en la ejecución de castigos, consolida la posición del BCP para gestionar eficientemente las fluctuaciones del ciclo macroeconómico (Breed, 2020).

Capítulo 6: Evaluación econométrica del impacto de la facturación electrónica de la SUNAT: Modelos, variables y heterogeneidad sectorial

La digitalización de la administración tributaria en el Perú, impulsada por la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT), constituye una de las transformaciones estructurales de la política fiscal más relevantes de la región. Mediante el despliegue progresivo del Sistema de Emisión Electrónica (SEE) —que integra la modalidad SOL (Portal), los sistemas informáticos propios del contribuyente y la intermediación de los Operadores de Servicios Electrónicos (OSE)—, la autoridad tributaria modificó sustancialmente el esquema de verificación e inspección del Impuesto General a las Ventas (IGV).

La cuantificación precisa de los efectos causales de esta reforma sobre el cumplimiento tributario, el reporte de operaciones gravadas y la recaudación efectiva ha requerido el diseño de complejos modelos econométricos de datos de panel. Estas metodologías permiten aislar el choque de política fiscal frente a variaciones macroeconómicas agregadas y tendencias dinámicas a nivel de empresa.

Estrategia empírica y especificación del modelo econométrico

Desde la perspectiva de la teoría de cumplimiento tributario, la implementación de la facturación electrónica (FE) reduce de manera drástica las asimetrías de información entre los agentes económicos y

la administración fiscal. Al estandarizar las comprobaciones de pago mediante formatos digitales estructurados en XML y autenticados con firma digital, el costo marginal de monitoreo para la SUNAT disminuye, elevando la probabilidad percibida de detección de la evasión y la elusión fiscal (Santisteban, 2021).

Para medir este impacto sin incurrir en sesgos de selección endógena ni en simultaneidad, la literatura econométrica especializada —particularmente los estudios elaborados por investigadores del Fondo Monetario Internacional y la SUNAT (Bellon et al., 2019, 2022)— explota la variabilidad cuasiexperimental del calendario de obligatoriedad escalonada por grupos o "olas" (*waves*) definido normativamente por la SUNAT.

Modelo de diferencias en diferencias escalonado con efectos fijos

La estrategia cuantitativa predominante se basa en un modelo de panel de Diferencias en Diferencias (DiD) con efectos fijos y tiempos de tratamiento heterogéneos. La especificación lineal se formaliza mediante la siguiente ecuación estructural:

$$Y_{it} = \alpha_i + \delta_t + \eta_i t + \beta \cdot I(\text{Treat}_{it} = 1) + \gamma X_{it} + \varepsilon_{it}$$

En esta formulación, Y_{it} denota la variable de resultado fiscal o contable expresada en soles corrientes para la empresa i en el periodo t. Para controlar la heterogeneidad inobservable que permanece constante a lo largo del tiempo —como la capacidad directiva, el sector de origen o la ubicación geográfica—, se incorpora el vector de efectos fijos individuales por empresa α_i. Asimismo, el vector de efectos fijos temporales δ_t absorbe los choques macroeconómicos comunes a

todas las firmas en un periodo determinado, como ciclos de inflación o fluctuaciones agregadas del producto.

Para relajar el supuesto de tendencias paralelas pretratamiento, el modelo introduce una tendencia temporal individual por firma $\eta_i t$, garantizando que las diferencias en las trayectorias de crecimiento previas a la reforma no contaminen el parámetro de interés. El término de tratamiento principal está representado por la variable dicotómica $I(\text{Treat}_{it} = 1)$, la cual toma el valor de 1 a partir del periodo en que la empresa i queda sujeta obligatoriamente al uso de la facturación electrónica y durante los periodos posteriores, y 0 en caso contrario. Finalmente, X_{it} es un vector de controles variantes en el tiempo, y ε_{it} representa el término de error, cuyos errores estándar se agrupan (*clustered*) a nivel de empresa para corregir problemas de correlación serial.

El horizonte muestral habitualmente analizado abarca observaciones trimestrales (por ejemplo, el periodo 2016T1–2023T2) sobre paneles balanceados compuestos por aproximadamente 200.000 pequeñas y microempresas con ventas anuales inferiores a 350 Unidades Impositivas Tributarias (UIT). Adicionalmente, los modelos incorporan reestimaciones sobre paneles no balanceados para examinar la dinámica de entrada y salida (*mortalidad empresarial*) generada por el costo de adopción.

Modelo de mínimos cuadrados en dos etapas (2SLS / variable instrumental)

La existencia de adopción voluntaria anticipada y de incumplimiento temporal de las fechas límite introduce un sesgo de endogeneidad en la estimación del efecto real del uso de la

herramienta. Para abordar esta limitación y distinguir entre la **Intención de Tratamiento** (*Intent-to-Treat* - ITT) y el **Efecto Promedio en los Tratados** (*Treatment on the Treated* - TOT), se especifica un esquema de Mínimos Cuadrados en Dos Etapas (2SLS):

$$\text{Primera Etapa:} \quad A_{it} = \alpha_i + \theta \cdot I(\text{Treat}_{it} = 1) + \gamma X_{it} + \delta_t + \eta_i t + \epsilon_{it}$$

$$\text{Segunda Etapa:} \quad Y_{it} = \alpha_i + \beta \hat{A}_{it} + \gamma X_{it} + \delta_t + \eta_i t + \varepsilon_{it}$$

En este esquema, la variable de adopción efectiva A_{it} es instrumentada mediante la asignación normativa exógena de la SUNAT $I(\text{Treat}_{it} = 1)$, lo que permite aislar la variación exógena en el cumplimiento y estimar el impacto directo de la emisión digital sobre los indicadores tributarios.

Operacionalización de variables y mecanismos de atenuación fiscal

El análisis econométrico evalúa el impacto de la reforma organizando los datos administrativos de la SUNAT en variables de resultado, variables explicativas de tratamiento y vectores de control de fricción operativa (véase la tabla 15).

Tabla 15: *Matriz estructural de variables econométricas*

Categoría de Variable	Variable Específica	Indicador / Unidad de Medida	Papel Econométrico	Efecto Empírico Estimado y Observaciones
Dependiente	Ventas Gravables (Y^{sales})	Soles corrientes (transformación logarítmica o niveles)	Variable de resultado principal	**Incremento de +6.5% a +7.4%** en el primer año, escalando hasta **+13% a +15%** al cuarto trimestre

Categoría de Variable	Variable Específica	Indicador / Unidad de Medida	Papel Econométrico	Efecto Empírico Estimado y Observaciones
				post-adopción.
Dependiente	Compras Gravables (Y^{purch})	Soles corrientes	Variable de resultado intermedia	**Incremento de +5.5%**. Refleja la exigencia de facturas electrónicas a proveedores formalizados.
Dependiente	Pago efectivo del IGV (Y^{VAT})	Soles recaudados en efectivo por la SUNAT	Variable de interés fiscal final	**Impacto positivo de +6.5% a +13%**, moderado temporalmente por compensación de saldos.
Dependiente	Empleo Registrado (Y^{emp})	Número de trabajadores en planilla electrónica	Variable de derrame formalizador	**Incremento moderado (~2%)**, sugiriendo formalización laboral inducida por la trazabilidad de operaciones.
Independiente	Mandato Legal ($I(\text{Treat})$)	Variable dummy (1 si $t \geq t_{\text{mandato}}$, 0 si no)	Tratamiento cuasiexperimental	Mide el choque normativo exógeno por grupo de obligatoriedad según resolución de SUNAT.

Categoría de Variable	Variable Específica	Indicador / Unidad de Medida	Papel Econométrico	Efecto Empírico Estimado y Observaciones
Independiente	Adopción Efectiva (A_{it})	Variable dummy o % de comprobantes emitidos vía SEE	Variable endógena (en 2SLS)	Captura la intensidad real de uso del sistema informático de facturación.
Control	Stock de Capital (K_{it})	Activos fijos declarados en soles	Control de escala productiva	Captura la capacidad instalada y escala operativa de la firma.
Control	Masa Salarial (W_{it})	Gastos de personal registrados en soles	Control de insumo laboral	Controla variaciones operativas internas no asociadas a la reforma fiscal.
Fricción Fiscal	Crédito Fiscal Acumulado	Saldo a favor del contribuyente ($t-1$)	Variable moderadora de liquidez	**Efecto Amortiguador**: Reduce la conversión directa de mayores ventas declaradas en recaudación de caja.

Mecanismo de transmisión y la fricción del crédito fiscal acumulado

Un hallazgo econométrico sustantivo en la evaluación del sistema peruano es la asimetría entre el incremento sustancial en las ventas gravables reportadas (+7.4% en la especificación ITT durante el

primer año) y la respuesta del pago efectivo de impuestos a la SUNAT. La causa subyacente de esta discrepancia radica en el **stock preexistente de créditos fiscales del IGV acumulados** por las empresas en periodos anteriores.

En la legislación tributaria peruana, el IGV pagado en las compras de insumos que supera al IGV generado por las ventas no es devuelto de forma inmediata en efectivo, sino que se constituye como un saldo a favor acumulativo que se arrastra hacia los periodos fiscales siguientes. Al ser obligadas a adoptar la facturación electrónica y registrar mayores ventas tributables, las empresas redujeron el impacto en su flujo de caja liquidando sus stocks de créditos fiscales acumulados en lugar de incrementar directamente sus pagos monetarios (Santisteban, 2021).

La dinámica del flujo monetario hacia la administración fiscal se representa mediante la relación de balances:

$$\begin{aligned}\Delta\text{Recaudación Efectiva}_{it} \\ &= \Delta\text{Débito Fiscal}_{it}(\text{Ventas}) \\ &- [\Delta\text{Crédito Fiscal}_{it}(\text{Compras}) \\ &+ \text{Uso de Saldos Acumulados}_{it-1}]\end{aligned}$$

Esta interacción evidencia que la digitalización como herramienta de control incrementa de forma inmediata la base imponible declarada, pero su conversión en caja fiscal depende de la estructura preexistente de créditos y de los mecanismos de devolución del sistema tributario.

La magnitud de la respuesta empresarial ante la adopción del sistema digital varía considerablemente según la estructura económica y productiva del sector donde opera la firma. Para capturar

esta variación, las estimaciones econométricas desagregan la muestra mediante regresiones por submuestras o incorporan términos de interacción sectoriales expresados en la especificación:

$$Y_{it} = \alpha_i + \delta_t + \beta_0 \cdot I(\text{Treat}_{it} = 1) + \sum_s \beta_s \left(I(\text{Treat}_{it} = 1) \times \text{Sector}_{i,s} \right) + \gamma X_{it} + \varepsilon_{it}$$

Donde $\text{Sector}_{i,s}$ representa una variable dicotómica para cada rama de actividad según la Clasificación Industrial Internacional Uniforme (CIIU) (véase la tabla 16).

Tabla 16: Análisis de heterogeneidad sectorial y efectos de red

Sector Económico (CIIU)	Riesgo de Evasión Pre-Reforma	Impacto Relativo en Ventas / IGV	Principal Mecanismo Econométrico Explicativo
Alojamiento y Servicios de Comida	Muy Alto	Muy Alto (Positivo)	La obligatoriedad redujo drásticamente la omisión de comprobantes en el canal final. Alta sensibilidad al control en tiempo real.
Transporte y Almacenamiento	Alto	Alto (Positivo)	Formalización en servicios de carga y pasaje. Mayor efectividad en la trazabilidad de costos operativos y de combustible.
Comercio al por Mayor y Menor	Moderado - Alto	Moderado (Positivo)	Dinámica mixta. En el segmento B2B, la demanda de crédito fiscal por compradores aceleró la formalización.
Manufactura	Bajo - Moderado	Moderado / Estable	Elevado cumplimiento previo. El impacto se concentró en la desarticulación de proveedores ficticios (*ONR*).

Sector Económico (CIIU)	Riesgo de Evasión Pre-Reforma	Impacto Relativo en Ventas / IGV	Principal Mecanismo Econométrico Explicativo
Construcción	Alto	Heterogéneo / Sensible	Alta rotación de subcontratistas. Los efectos se asociaron al control cruzado de deducciones por insumos y materiales.

Factores explicativos de la heterogeneidad sectorial

Primero, el grado de informalidad estructural previo a la reforma determina la escala del margen intensivo y extensivo de formalización. En actividades caracterizadas por un elevado uso de efectivo y baja bancarización, como los servicios de comida, hotelería y transporte terrestre, la fiscalización trazable en tiempo real generó los mayores incrementos relativos en la declaración de operaciones gravadas.

Segundo, la posición de las firmas en la cadena de valor (B2B frente a B2C) altera los incentivos intertemporales. En las transacciones entre empresas (B2B), el cliente exige la emisión de una factura electrónica válida para sustentar crédito fiscal del IGV y gasto deducible del Impuesto a la Renta, generando un mecanismo de autofiscalización cruzada. Por el contrario, en las transacciones con consumidores finales (B2C), la ausencia de un incentivo directo al comprador requiere controles complementarios para evitar la subdeclaración de ingresos (Castillo y Lahura, 2022).

Tercero, la madurez tecnológica y el tamaño de la empresa actúan como barrera o facilitador de la transición. Mientras que las medianas y grandes empresas absorben la inversión informática mediante ganancias de eficiencia y simplificación contable, las

microempresas enfrentan fricciones que en algunos casos incrementaron las tasas de salida del mercado en las fases previas al cumplimiento obligatorio.

Efectos de desbordamiento en redes interempresariales (*Spillovers*)

Un aporte decisivo de la investigación econométrica aplicada al caso peruano (Bellon et al., 2022) consiste en el análisis de los **efectos de propagación o red (*spillovers*)** a través de las cadenas de valor. Utilizando las matrices de transacciones empresa a empresa construidas a partir de la información contable, los investigadores evaluaron cómo el mandato asignado a una empresa influye en el comportamiento de sus socios comerciales no sujetos a la norma.

Los resultados confirman que la adopción obligatoria de la facturación electrónica por parte de firmas grandes genera incentivos para que sus proveedores y clientes no obligados adopten voluntariamente el sistema de emisión electrónica. Este canal de transmisión es más pronunciado en sentido descendente (*downstream*), impulsado por la exigencia de comprobantes digitales por parte de las empresas compradoras (Santisteban, 2021).

Asimismo, los modelos de panel identifican efectos indirectos significativos sobre el cumplimiento tributario de las micro y pequeñas empresas. Específicamente, las firmas de menor tamaño no obligadas que mantienen relaciones comerciales con socios forzados a ingresar al sistema **reportan en promedio un 11% más de ventas y incrementan sus pagos de IGV en un 17%** durante el año en que su socio comercial adopta la facturación electrónica. Este fenómeno demuestra la existencia de economías de red en la fiscalización digital, donde el

control ejercido sobre un nodo central de la cadena productiva se propaga hacia los eslabones periféricos, expandiendo la base tributaria efectiva sin necesidad de intervención directa sobre cada unidad económica.

No obstante, la evidencia también advierte sobre dinámicas de segmentación de mercado. En sectores con elevada informalidad, la imposición de la facturación electrónica provocó que algunas firmas no formalizadas discontinuaran sus relaciones comerciales con socios obligados para reorientar sus ventas hacia circuitos comerciales puramente informales, evitando la trazabilidad fiscal (Castillo y Lahura, 2022).

Hallazgos empíricos consolidados y robustez econométrica

- Las estimaciones econométricas aplicadas sobre la base de datos de panel de la SUNAT ofrecen parámetros sólidos sobre la magnitud cuantitativa del impacto de la reforma (Santisteban, 2021):

- **Efecto promedio sobre ventas registradas**: En las especificaciones de Diferencias en Diferencias sobre el panel balanceado para el periodo 2017–2023, la facturación electrónica genera un incremento promedio del **6.5%** en los indicadores de desempeño impositivo. En el corto plazo, los modelos de Intención de Tratamiento (ITT) señalan un aumento del **7,4%** en las ventas gravables declaradas durante el primer año de obligatoriedad, acelerándose hasta alcanzar un nivel superior al **13%** hacia el cuarto trimestre posterior a la fecha límite de adopción.

- **Efecto promedio sobre compras registradas**: El volumen de compras declaradas se eleva en promedio un **5,5%**. Este resultado confirma que el aumento en el reporte de ingresos no fue contrarrestado por una generación fraudulenta de compras o por la inclusión masiva de facturas falsas.
- **Ajustes en la supervivencia y formalidad laboral**: Las estimaciones realizadas sobre paneles no balanceados detectan un incremento estadísticamente significativo en la probabilidad de salida del padrón activo para firmas en el tramo inferior de ingresos durante el periodo de anuncio de la norma, sugiriendo la salida de unidades económicas cuya viabilidad dependía de la omisión fiscal. En el plano laboral, la mayor trazabilidad de las operaciones comerciales generó un incremento secundario cercano al **2%** en el número de trabajadores declarados en la planilla electrónica, acompañando el incremento del valor agregado reportado.

Las pruebas de robustez econométrica confirman la consistencia de estos estimadores. Las estimaciones con tendencias específicas por firma confirman la validez del supuesto de tendencias paralelas en los periodos de pretratamiento. Por su parte, la comparación entre las especificaciones ITT y TOT mediante variables instrumentales evidencia que el sesgo por adopción voluntaria o retraso normativo no altera la dirección ni la significancia estadística de los coeficientes centrales.

La evaluación econométrica del impacto de la facturación electrónica en el Perú confirma que la digitalización de los registros comerciales constituye un instrumento de alta efectividad para fortalecer el cumplimiento tributario y reducir la evasión en el IGV. Las

metodologías de Diferencias en Diferencias con adopción escalonada permiten establecer con solidez causal que la disminución de los costos de verificación para la SUNAT induce un mayor reporte de ventas y compras gravadas (Medina, 2023).

Sin embargo, los resultados empíricos demuestran que la efectividad fiscal de las tecnologías de la información no es independiente de la arquitectura normativa e institucional del sistema tributario. La acumulación previa de créditos fiscales frena la conversión directa del mayor reporte de operaciones en recaudación monetaria de corto plazo, subrayando la importancia de acompañar las reformas de fiscalización digital con un rediseño de los esquemas de compensación y devolución de impuestos.

Asimismo, la presencia de amplios efectos de red (*spillovers*) a través de las cadenas B2B implica que la autoridad tributaria puede maximizar el impacto de sus intervenciones priorizando la fiscalización de nodos empresariales estratégicos y reduciendo las barreras tecnológicas para las micro y pequeñas empresas, evitando así la fragmentación del mercado hacia circuitos informales.

Conclusión

La elaboración de este libro de investigación demuestra que el análisis de la realidad económica peruana exige trascender las fronteras del modelado unidimensional y la dependencia exclusiva de las cuentas nacionales agregadas. A través del marco conceptual, metodológico y empírico desarrollado a lo largo de esta obra, se consolida la integración analítica de las tres fuentes primarias más relevantes del país: el registro fiscal tributario (SUNAT), la transaccionalidad bancaria y crediticia privada (BCP) y los agregados estadísticos y socioeconómicos oficiales (INEI).

¿De qué manera la integración de datos tributarios (SUNAT), financieros (BCP) y macroeconómicos (INEI) mediante modelos econométricos y estadísticos avanzados permite capturar la heterogeneidad de la economía peruana, optimizar la precisión de los pronósticos y mejorar la toma de decisiones públicas y privadas? La integración analítica de estas tres fuentes opera como una arquitectura tripartita que complementa la oportunidad temporal, la granularidad microeconómica y la representatividad macroeconómica.

- Captura de la heterogeneidad económica: Mientras que las estadísticas del INEI ofrecen la cobertura representativa nacional, la inclusión de los microdatos de la SUNAT y la transaccionalidad del BCP permiten mapear las diferencias regionales y sectoriales en tiempo cercano al real (nowcasting). Esta combinación reduce los sesgos generados por la alta informalidad estructural, permitiendo observar patrones diferenciados entre la economía formal consolidada, los

segmentos microempresariales transicionantes y el consumo de las familias no bancarizadas.

- Optimización del poder predictivo: Los modelos integrados (VEC, datos de panel dinámicos y algoritmos de machine learning como XGBoost y Random Forest) superan sistemáticamente a los enfoques univariados y tradicionales basados en una sola fuente. El error cuadrático medio (RMSE) en las proyecciones del PBI trimestral y del consumo privado se reduce de manera significativa al incorporar la facturación electrónica de la SUNAT y la transaccionalidad del BCP como variables explicativas y de señalización anticipada.
- Mejora en la toma de decisiones:

• Proporciona un instrumental riguroso para la formulación de políticas tributarias más equitativas, la evaluación de la eficacia de los incentivos fiscales regionales y la simulación del impacto macroeconómico de choques exógenos en la recaudación y el empleo.

• Ofrece a las entidades financieras y a la alta dirección corporativa un marco superior para la estimación de demanda sectorial, la gestión del riesgo de crédito por regiones y la planificación de inversiones basada en indicadores oportunos de liquidez y mercado.

Ahora bien, la adopción generalizada de los Comprobantes de Pago Electrónicos (CPE) ha transformado la facturación en un termómetro de alta frecuencia de la actividad real. Las variaciones en las declaraciones mensuales de ventas no solo correlacionan fuertemente con el Valor Agregado Bruto sectorial del INEI, sino que

reducen el desfase de estimación de 60 días (propio de los indicadores macroeconómicos habituales) a pocos días del cierre del período fiscal.

Mediante el modelado de brechas entre el volumen transaccional bancario, las ventas declaradas a la SUNAT y las estimaciones de empleo e ingresos de la ENAHO (INEI), se identificaron patrones de subdeclaración e informalidad con un alto grado de especificidad geográfica y de cadena de valor (destacando el comercio minorista, el transporte y la construcción informal). La integración de estas métricas permite ajustar las cifras del PBI regional con mayor precisión y diseñar estrategias tributarias dirigidas sin asfixiar a las micro y pequeñas empresas (MYPE).

Por lo que la evidencia empírica confirma que el volumen de facturación a través de canales digitales y adquirencia (POS, transferencias y billeteras digitales), así como la evolución de los saldos en cuentas corrientes y de ahorros del BCP, anteceden entre 1 y 2 meses a las cifras oficiales de consumo privado publicadas por el INEI. A su vez, el ritmo de colocación y refinanciación de la cartera de crédito empresarial y de consumo del BCP actúa como un predictor altamente sensible de la Inversión Bruta Fija privada. Los incrementos en la morosidad o la ralentización de las líneas de capital de trabajo funcionan como señales tempranas de desaceleración económica sectorial, permitiendo construir índices de tensión financiera regional antes de que los efectos se traduzcan en contracciones del empleo o caída del PBI.

En este sentido, las técnicas de series de tiempo multivariadas (VAR/VEC) demostraron ser efectivas para capturar las interacciones de largo plazo entre el Índice de Precios al Consumidor (IPC), las tasas de interés y el PBI. Sin embargo, cuando estas metodologías

tradicionales se combinan con modelos de datos de panel y algoritmos de machine learning, la capacidad analítica se multiplica. Y la aplicación de algoritmos de machine learning sobre las bases unificadas permitió identificar relaciones no lineales y de interacción compleja entre la inflación (INEI), la capacidad de pago y mora bancaria (BCP) y la carga tributaria efectiva (SUNAT). Los modelos híbridos resultaron especialmente superiores en la segmentación del riesgo crediticio y la predicción de quiebras o ceses de operaciones de empresas en entornos de alta volatilidad macroeconómica.

En conclusión, el futuro de la investigación aplicada y la consultoría económica en el Perú radica en el uso coordinado de los datos masivos financieros y fiscales junto con la rigurosidad metodológica de la estadística oficial. La integración de los datos de la SUNAT, el BCP y el INEI no solo proporciona una radiografía más limpia y fidedigna de la estructura socioeconómica peruana, sino que democratiza e impulsa el desarrollo de un ecosistema analítico moderno para afrontar con éxito los retos del desarrollo nacional.

Bibliografía

Balmore, A.R. (2023). Análisis de las elasticidades de la recaudación tributaria en El Salvador. *Empresa y Sociedad,* 4, 89-100

Breed, DG, Hurter, J., Marimo, M., Raletjene, M., Raubenheimer, H., Tomar, V., & Verster, T. (2023). Una metodología prospectiva de la NIIF 9, centrada en la incorporación de información macroeconómica y macroprudencial en el cálculo de las pérdidas crediticias esperadas. *Risks* , *11* (3), 59. https://doi.org/10.3390/risks11030059

Brum, C. Y Rodríguez, H. (2016). *Modelos puente para proyectar el PIB en el corto plazo. Enfoque sectorial.* Montevideo: Banco Central del Uruguay, Depto de Análisis de Coyuntura, Área de Análisis Macroeconómico. https://www.bcu.gub.uy/Estadisticas-e-Indicadores/Documentos%20de%20Trabajo/10.2016.pdf

Castillo, T., y Lahura, E. (2022). Los Efectos de la Política Fiscal sobre la Actividad Económica en el Perú. *Revista Estudios Económicos,* 40, 9 – 34. https://www.bcrp.gob.pe/docs/Publicaciones/Revista-Estudios-Economicos/40/ree-40-castillo-lahura.pdf

Conejo, C., Otoya, M., y Cardoza, D. (2011). INGRESOS FISCALES Y ELASTICIDADES TRIBUTARIAS:ESTIMACIÓN DE LAS ELASTICIDADES TRIBUTARIAS DE CORTO Y LARGO PLAZO PARA LOS PRINCIPALES IMPUESTOS. Economía y Sociedad, 39/40, 47-61

Coroneo, L., Giannone, D., & Modugno, M. (2016). Unspanned Macroeconomic Factors in the Yield Curve. *Journal of Business & Economic Statistics*, *34*(3), 472–485. https://doi.org/10.1080/07350015.2015.1052456

El Comercio. (2024, 12 junio). BCP: transacciones en el ecosistema de consumo masivo crecieron más de 30% en el último año. https://elcomercio.pe/economia/peru/bcp-transacciones-en-el-ecosistema-de-consumo-masivo-crecieron-mas-de-30-en-el-ultimo-ano-i-ultimas-noticia/?ref=ecr

Ervina, N., Junaidi, J., Zulgani, Z., & Achmad, E. (2025). Analysis of Tax Compliance Levels for Regional Taxes in the Provinces of Indonesia. *Economies*, *13*(12), 354. https://doi.org/10.3390/economies13120354

Espinosa-Zúñiga, J.J.,(2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y Tecnología*, 21 (03), 1-16. https://doi.org/10.22201/fi.25940732e.2020.21.3.022

Gordón-Mendoza, R., & Camargo-Buitrago, I. (2015). Selección de estadísticos para la estimación de la precisión experimental en ensayos de maíz. *Agronomía Mesoamericana*, *26*(1), 56-63

Greenwood, C. J., Youssef, G. J., Letcher, P., Macdonald, J. A., Hagg, L. J., Sanson, A., Mcintosh, J., Hutchinson, D. M., Toumbourou, J. W., Fuller-Tyszkiewicz, M., & Olsson, C. A. (2020). A comparison of penalised regression methods for informing the selection of predictive markers. *PloS one*, *15*(11), e0242730. https://doi.org/10.1371/journal.pone.0242730

Hendry, D., & Ericcson, N. (2003). *Understanding Economic Forecasts.* Cambridge: The MIT press.

Higgins, Patrick C., GDPNow: A Model for GDP 'Nowcasting' (July 1, 2014). FRB Atlanta Working Paper 2014-2017. http://dx.doi.org/10.2139/ssrn.2580350

INEI. (2002). *Control de calidad en los procesos estadísticos: una aproximación basada en los modelos ARIMA con análisis de Intervención*. Lima: Instituto Nacional de Estadística e Informática. https://cdn.www.gob.pe/uploads/document/file/4093175/Control%20de%20Calidad%20en%20los%20Procesos%20Estad%C3%ADsticos.pdf?v=1675703147

INEI. (2017). *Evaluación de la Calidad del Censo de Población 2017. Síntesis Metodológica Nº 7*. Lima: Instituto Nacional de Estadística e Informática. https://cdn.www.gob.pe/uploads/document/file/2472024/Evaluaci%C3%B3n%20de%20la%20Calidad%20del%20Censo%20de%20Poblaci%C3%B3n%202017.%20S%C3%ADntesis%20Metodol%C3%B3gica%20N%C2%BO%207.pdf?v=1637371198

INEI. (2017). *Metodología de la Estimación de la Población Económicamente Activa Ocupada.* Lima: Instituto Nacional de Estadística e Informática. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1676/01.pdf

INEI. (2021). *Metodología del índice de precios al consumidor a nivel nacional base diciembre 2021*. Lima: Instituto Nacional de Estadística e Informática. https://cdn.www.gob.pe/uploads/document/file/4097881/Metodolog%C3%ADa%20del%20%C3%8Dndice%20de%20Precios%20al%20Consumidor%20a%20Nivel%20Nacional%20-%20Base%20Diciembre%202021.pdf?v=1784734041

INEI. (2025). *Metodología para el cálculo de los índices unificados de precios de la construcción (Base Diciembre 2025)*. Lima: Instituto

Nacional de Estadística e Informática. https://cdn.www.gob.pe/uploads/document/file/10354686/2439722-metodologia-para-el-calculo-de-los-indices-unificados-de-precios-de-la-construccion-base-diciembre-2025.pdf?v=1784734041

Instituto Nacional de Estadística e Informática. (1999). *Resumen metodológico del cambio de año base de las cuentas nacionales del Perú*. https://proyectos.inei.gob.pe/web/biblioineipub/bancopub/Est/Lib0376/indice.htm

Luo, J., & Yu, C. L. (2021). Determining Number of Factors in Dynamic Factor Models Contributing to GDP Nowcasting. *Mathematics,* 9(22), 2865. https://doi.org/10.3390/math9222865

Medina, E. (2023). *Evasión Tributaria del IGV y Renta con Incidencia en los Estados Financieros de una Empresa de Servicios, Lima - periodo 2016 -2019* (Tesis de Grado, Universidad Ricardo Palma). Repositorio Institucional – Universidad Ricardo Palma. https://repositorio.urp.edu.pe/server/api/core/bitstreams/584a2bd4-2385-4052-9023-4c454095c3ca/content

Mendoza-Velarde, P. G. (2025). Relación de largo y corto plazo entre la recaudación de impuestos y el crecimiento económico en el Perú: evidencia de cointegración y causalidad. *Quipukamayoc*, *33*(70), 9-23. https://doi.org/10.15381/quipu.v33i70.31246

Pelegrí Álvarez, V. (2019). *El filtro de Kalman* (Tesis de Grado, Universitat de Barcelona). Repositorio Institucional – Universitat de Barcelona. https://diposit.ub.edu/server/api/core/bitstreams/2ca6623a-af07-49d2-93c9-1413182379e4/content

Pérez Forero, F. (2021). "Predicción de variables macroeconómicas en el Perú a través un modelo BVAR con media cambiante en el tiempo," Working Papers 2021-001, Banco Central de Reserva del Perú. https://ideas.repec.org/p/rbp/wpaper/2021-001.html

Pérez-Forero, F.J. (2020). Proyecciones en tiempo Real del PBI utilizando Indicadores Líderes y elección Bayesiana de variables. *Revista BCRP,* 1-19. https://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2018/documento-de-trabajo-010-2018-esp.pdf

Santisteban Morales, J.E. (2021). *Implementación de la facturación electrónica y su impacto en el impuesto general a las ventas e impuesto a la renta en una empresa de venta de alimentos balanceados para la ganadería del distrito de Lurín periodo 2017-2018* (Tesis de Grado, Universidad Ricardo Palma). Repositorio Institucional – Universidad Ricardo Palma. https://hdl.handle.net/20.500.14138/4274

Silva Romero, M.A. (2013). Modelo de pronóstico para la estimación de la utilización y confiabilidad de equipos dinámicos. Caso: Equipo de compresión de la RPMNE. Ingeniería Petrolera, 53(5), 287-297. https://biblat.unam.mx/hevila/Ingenieriapetrolera/2013/vol53/no5/3.pdf

Tapia Romero, M. A. (2025). *Presión fiscal y crecimiento económico: Análisis comparativo entre países de la CAN* (Tesis de maestría, Universidad Técnica de Ambato). Repositorio Institucional UTA. https://repositorio.uta.edu.ec/handle/123456789/45629

Vásquez Trujillo, M. C. (2023). Dinámica de las elasticidades tributarias: análisis para Guatemala 2008-2020. Revista Científica del Sistema de

Estudios de Postgrado. 6(1). 73-86. DOI: https://doi.org/10.36958/sep.v6i1.153

Vásquez Trujillo, M.C. (2023). Dinámica de las elasticidades tributarias: análisis para Guatemala 2008-2020. *Revista Científica del Sistema de Estudios de Postgrado.* 6(1). 73-86. DOI: https://doi.org/10.36958/sep.v6i1.153

Ventura Hernández, E. M., Guzmán Castro, E. A., Matta Solis, E. P., Muñoz Chacabana, J. L., Vega Diaz, E., Condori Uria, R. A. S., & Hinostroza Atahualpa, A. A. (2024). La evasión tributaria y su impacto en la recaudación fiscal en Perú y América Latina. Zenodo. https://doi.org/10.5281/zenodo.10811143

Weller, J. (2014). Aspectos de la evolución reciente de los mercados laborales de América Latina y el Caribe. *Revista CEPAL,* 114, 7-30. https://repositorio.cepal.org/server/api/core/bitstreams/0da64014-403a-45da-89d2-f11eb45f1ae6/content

Winkelried, D. y Urquizo, C. (2021. The Economic Effects of International Administrations: The Cases of Kosovo and East Timor. Economic Development and Cultural Change, University of Chicago Press, vol. 69(2), pages 869-901

Zou, H., y Trevor, H. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B.,* 67(2), 301–320. https://sites.stat.washington.edu/courses/stat527/s13/readings/zouhastie05.pdf

Esta edición de *"Modelos econométricos y estadísticos basados en datos de la SUNAT, el BCRP y el INEI"* fue finalizada en la ciudad de Montevideo, en la República Oriental del Uruguay, el 22 de septiembre de 2026.

AUSTRAL

MODELOS ECONOMÉTRICOS Y ESTADÍSTICOS BASADOS EN DATOS DE LA SUNAT, EL BCRP Y EL INEI

WWW.EDITORIALAUSTRAL.COM

ISBN: 978-9915-720-05-0