Modelo GAMLSS–BEINF para la recuperación de créditos de consumo respaldados por garantías

Oscar Steven Martinez Ardila

Trabajo de Grado presentado para optar al título de Especialista en Estadística

Director

Heivar Yesid Rodríguez Pinzón

Magister en Ciencias Económicas

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemáticas

Especialización en Estadística

Bucaramanga

2025

Dedicatoria

A mi esposa Nydia Milena y mi hija Maria Paula, por su amor incondicional, paciencia y total apoyo en cada etapa de este proceso. Ustedes fueron mi inspiración constante y son el corazón de cada logro que alcanzo.

A mi mamá Luz Elena y a mi papá Abel, por su amor permanente y su confianza constante en mis decisiones. Gracias por haber sembrado en mí los valores de perseverancia y dedicación que me han llevado hasta aquí. Este logro es reflejo de la fortaleza que ustedes me inculcaron desde siempre.

Dedico este trabajo a todos ustedes, reconociendo que su amor, inspiración y fe en mí han sido el pilar fundamental para la culminación de esta especialización.

Agradecimientos

Agradezco a mi director de grado, por su acompañamiento, orientación y dedicación en cada etapa de este proceso.

Agradezco a cada uno de los profesores de la Especialización en Estadística por enseñarme cómo aplicar la estadística en sus diferentes formas y contribuir con sus conocimientos al desarrollo de este trabajo.

Finalmente, agradezco a la Universidad Industrial de Santander por facilitar los recursos y espacios necesarios para la realización de esta especialización.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Resumen

Título: Modelo GAMLSS–BEINF para la recuperación de créditos de consumo respaldados por garantías*

Autor: Oscar Steven Martínez Ardila**

Palabras clave: Recuperación de créditos, Riesgo crediticio, Distribución Beta Inflada (BEINF), GAMLSS, Regresión distribucional, Garantías financieras, Pérdida dada el incumplimiento

Descripción:

Este estudio desarrolla un modelo predictivo basado en la distribución Beta Inflada en Ceros y Unos (BEINF) dentro del marco GAMLSS para analizar la recuperación de créditos de consumo siniestrados respaldados por garantías. El análisis abarca 11.031 operaciones crediticias durante 2013-2023, periodo en el cual la variable de recuperación exhibe una estructura compleja caracterizada por inflación en los extremos: 55,9% de casos sin recuperación (R=0), 14,3% con recuperación completa (R=1) y 29,8% con recuperación parcial. Esta particular estructura distribucional justifica la aplicación de la metodología BEINF, que permite modelar simultáneamente la probabilidad de los extremos y la conducta de la recuperación intermedia.

El modelo especifica cuatro componentes distribucionales para capturar mecanismos distintos de recuperación: μ (media de recuperación parcial), σ (dispersión), ν (probabilidad de pérdida total) y τ (probabilidad de recuperación completa). La selección de variables se realiza mediante procedimientos stepwise independientes para cada parámetro, identificando factores predictivos relevantes. De esta manera, se proporciona una herramienta predictiva robusta que mejora la toma de decisiones estratégicas y la eficiencia operativa de la gestión de cobranzas.

El modelo aporta una metodología para la recuperación de cartera dado que identifica que el primer abono constituye un punto de quiebre crítico en la trayectoria de recuperación, permitiendo priorizar estrategias de contacto temprano y negociación inicial. Las implicaciones operacionales sugieren focalizar recursos en los primeros meses post-siniestro, reconociendo que la recuperación es un fenómeno multidimensional que demanda intervenciones específicas según el escenario de cada operación.

---

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Matemáticas. Director Heivar Yesid Rodríguez Pinzón.

Abstract

Title: GAMLSS–BEINF Model for Recovery of Consumer Credit Backed by Guarantees*

Author: Oscar Steven Martínez Ardila**

Keywords: Credit recovery, Credit risk, Beta Inflated Distribution (BEINF), GAMLSS,

Distributional regression, Financial guarantees, Loss given default

Description:

This study develops a predictive model based on the Zero-and-One Inflated Beta Distribution (BEINF) within the GAMLSS framework to analyze the recovery of defaulted consumer credit operations backed by guarantees. The analysis encompasses 11,031 credit operations during 2013-2023, a period in which the recovery variable exhibits a complex structure characterized by inflation at the extremes: 55.9% of cases with zero recovery (R=0), 14.3% with complete recovery (R=1), and 29.8% with partial recovery. This particular distributional structure justifies the application of the BEINF methodology, which enables simultaneous modeling of the probability of extremes and the behavior of intermediate recovery.

The model specifies four distributional components to capture distinct recovery mechanisms: μ (mean partial recovery), σ (dispersion), ν (probability of total loss), and τ (probability of complete recovery). Variable selection is performed through independent stepwise procedures for each parameter, identifying relevant predictive factors. In this manner, a robust predictive tool is provided that improves strategic decision-making and operational efficiency of collection management.

The model contributes a methodology for portfolio recovery by identifying that the first payment constitutes a critical turning point in the recovery trajectory, enabling prioritization of early contact strategies and initial negotiation. The operational implications suggest focusing resources in the first month's post-default, acknowledging that recovery is a multidimensional phenomenon that demands specific interventions according to each operation's scenario.

---

* Bachelor Thesis

** Facultad de Ciencias. Escuela de Matemáticas. Director Heivar Yesid Rodríguez Pinzón.

Introducción

El crédito de consumo de libre inversión se ha consolidado como un motor fundamental para la economía, al permitir que individuos y familias financien diversas necesidades, que van desde educación y salud hasta la consolidación de deudas o la creación de nuevos emprendimientos. Sin embargo, su facilidad de acceso y flexibilidad también implican riesgos inherentes.

Uno de los desafíos más críticos para las entidades que otorgan este tipo de créditos es la gestión de la cartera incobrable, particularmente con mora superior a 180 días. Este tipo de cartera representa una pérdida directa de capital, afecta la liquidez y disminuye la rentabilidad de las instituciones, pudiendo comprometer su estabilidad financiera. Además, incrementa las provisiones requeridas y reduce la capacidad para otorgar nuevos créditos, lo que repercute negativamente en su sostenibilidad.

Como complemento a las estrategias de mitigación, los sistemas de garantías se han constituido en mecanismos clave dentro del sistema financiero para cubrir el riesgo de las operaciones crediticias, disminuyendo los problemas de asimetría de información entre prestamistas y prestatarios.

El presente trabajo de grado tiene como propósito desarrollar un modelo para estimar la tasa de recuperación de operaciones respaldadas por una entidad de garantías. Para este fin, se propone emplear la distribución beta inflada en cero y uno, la cual es una extensión de la distribución beta que permite modelar variables proporcionales definidas en el intervalo abierto (0,1), pero que también pueden asumir, con probabilidad positiva, los valores extremos exactos 0 y 1. Este enfoque es particularmente adecuado para variables como las tasas de recuperación, que

se encuentran naturalmente en el intervalo cerrado y que suelen presentar acumulaciones en los valores extremos, ya sea en casos de recuperación total (1) o sin recuperación (0).

El objetivo general de esta investigación es desarrollar una metodología predictiva que contribuya a mejorar las tasas de recuperación del Fondo de Garantías, favoreciendo una gestión más proactiva y eficiente de los recursos, así como la mitigación del impacto de las pérdidas monetarias.

1. Antecedentes

El capítulo de antecedentes se centra en la revisión de investigaciones previas relacionadas con el tema de estudio, tanto a nivel internacional como nacional.

1.1 Antecedentes Internacionales

Los primeros desarrollos teóricos sobre distribuciones beta infladas fueron presentados por Ospina y Ferrari (2010), quienes propusieron una extensión de la distribución beta clásica para abordar las limitaciones en el modelado de variables proporcionales que pueden asumir valores exactos en los extremos del intervalo [0,1]. Estos autores demostraron que la distribución beta inflada proporciona una mejor representación de datos proporcionales con acumulación en cero y uno, superando las restricciones de la distribución beta tradicional que no permite probabilidades positivas en los extremos del intervalo.

Hossain et al. (2016) desarrollaron metodologías para la estimación de percentiles en variables de respuesta proporcional utilizando el framework GAMLSS (Generalized Additive

Models for Location, Scale and Shape). Su trabajo demostró la aplicabilidad práctica de las distribuciones infladas en contextos de regresión, estableciendo las bases computacionales para el modelado flexible de todos los parámetros distribucionales como funciones de variables explicativas.

El desarrollo metodológico más completo fue proporcionado por Rigby et al. (2019) en su obra sobre distribuciones para modelado de localización, escala y forma usando GAMLSS en R. Los autores presentaron la implementación computacional completa de las distribuciones beta infladas, incluyendo las funciones de densidad, distribución acumulada y generación aleatoria, así como los algoritmos de estimación por máxima verosimilitud dentro del framework GAMLSS.

Stasinopoulos et al. (2024) en su tratado sobre modelos aditivos generalizados para localización, escala y forma, consolidaron el marco teórico y práctico de la regresión distribucional. Su trabajo proporciona una perspectiva integral sobre el modelado simultáneo de todos los parámetros de la distribución, estableciendo GAMLSS como una metodología estándar para análisis de datos complejos que no siguen distribuciones exponenciales clásicas.

En el contexto específico del riesgo crediticio, Altman et al. (2005) establecieron los fundamentos del modelado de tasas de recuperación, identificando los desafíos inherentes al análisis de variables proporcionales en el ámbito financiero. Su trabajo pionero demostró la necesidad de metodologías especializadas para el tratamiento de datos de recuperación con características distribucionales complejas.

Basel Committee on Banking Supervision (2017) estableció los marcos regulatorios para la medición del riesgo crediticio, definiendo la pérdida esperada como el producto de la probabilidad de incumplimiento, exposición al incumplimiento y pérdida dado el incumplimiento.

Esta formulación destacó la importancia crítica del modelado preciso de las tasas de recuperación en la gestión del riesgo financiero.

1.2 antecedentes nacionales

En el contexto colombiano, los estudios sobre modelado de recuperación crediticia han sido limitados, concentrándose principalmente en aspectos descriptivos y de gestión operativa. Galvis (2019) desarrolló un modelo para el cálculo de probabilidad de recuperación en el sector cooperativo, empleando metodologías tradicionales de regresión logística sin considerar la naturaleza específica de las variables proporcionales ni la inflación en extremos.

La Superintendencia de Economía Solidaria (2019) ha establecido marcos normativos para el cálculo del deterioro por riesgo de crédito, definiendo metodologías estándar basadas en la Pérdida Dado el Incumplimiento (PDI). Aunque estos marcos proporcionan directrices operativas importantes, no abordan las complejidades distribucionales inherentes a las tasas de recuperación ni aprovechan metodologías estadísticas avanzadas.

Los trabajos de modelado de riesgo crediticio en Colombia han seguido tradicionalmente enfoques basados en modelos lineales generalizados clásicos, sin explorar las ventajas de la regresión distribucional para variables con características específicas como las tasas de recuperación. Esta situación representa una oportunidad significativa para contribuir al campo mediante la aplicación de metodologías estadísticas avanzadas.

2. Justificación

La creación y consolidación de fondos de garantía ha transformado la dinámica crediticia para las micro, pequeñas y medianas empresas (mipymes), al funcionar como un mecanismo de respaldo que sustituye o complementa las garantías tradicionales exigidas por las entidades financieras. Gracias a estos fondos, los bancos comparten el riesgo de incumplimiento con un garante especializado, lo cual amplía la oferta de crédito sin incrementar la exposición de las instituciones prestamistas (CEPAL, 2021).

Desde el punto de vista operativo, los fondos de garantía ofrecen una activación ágil del aval: basta con la declaración de incumplimiento para que el garante cubra la pérdida, evitando procesos judiciales complejos de liquidación de colaterales reales y mejorando la eficiencia del sistema (CEPAL, 2021).

En el ámbito regulatorio, las garantías emitidas por los fondos reciben ponderaciones de menor riesgo bajo los estándares de Basilea III, lo que reduce los activos ponderados por riesgo y, en consecuencia, el capital mínimo exigido a los bancos. Este beneficio liberador de recursos promueve la concesión de nuevos créditos y optimiza los indicadores de rentabilidad (CASFOG, 2024; NAFIN, 2022).

Durante episodios de inestabilidad económica, los fondos de garantía han demostrado un enfoque contracíclico al ampliar temporalmente los niveles de cobertura, subsidiar comisiones y ajustar criterios de elegibilidad. Estas medidas mantienen activa la cadena de pagos y aseguran el capital de trabajo de las empresas, atenuando el sesgo procíclico de la banca tradicional (AECM, 2024).

No obstante, cuando las operaciones avaladas devienen en cartera siniestrada y la titularidad del crédito recae en el fondo, surgen retos significativos:

1. Asimetría de información con el deudor: al asumir la administración del crédito, el fondo carece de la relación previa con el prestatario, lo que dificulta la reanudación de contactos y la negociación de planes de pago (ALIDE, 2018).
2. Capacidad operativa limitada: la recuperación efectiva requiere equipos multidisciplinarios (abogados, valuadores, gestores de cobranza) y sistemas de seguimiento especializados, los cuales muchas veces resultan insuficientes ante volúmenes elevados de expediente (ALIDE, 2018).
3. Menor presión reputacional: sin la presencia directa del banco original, el prestatario percibe menos riesgo reputacional y puede demorar el cumplimiento de acuerdos de pago (CEPAL, 2021).

Para enfrentar estas limitaciones, en este trabajo se desarrollará y fortalecerá un modelo estadístico de distribución beta inflada, que permitirá identificar los factores determinantes de recuperación de cartera y mejorar la efectividad de la gestión de cobro mediante la cuantificación probabilística de resultados esperados.

3. Objetivos

3.1 Objetivo General

Desarrollar un modelo predictivo basado en la distribución beta inflada en cero y uno (BEINF) para la tasa de recuperación de los créditos respaldados por una entidad de garantías, con el fin de apoyar la toma de decisiones estratégicas y fortalecer la gestión del área de cobranzas.

3.2 Objetivos Específicos

- Analizar el comportamiento histórico de las tasas de recuperación de créditos respaldados por la entidad de garantías, identificando patrones y concentraciones en los valores extremos.
- Seleccionar, depurar y evaluar las variables explicativas que influyen en la probabilidad de recuperación total, parcial o nula.
- Ajustar modelos estadísticos utilizando la familia de distribuciones BEINF dentro del marco de GAMLSS, para capturar la presencia de inflación en 0 y 1 en la variable de recuperación.
- Comparar el desempeño del modelo propuesto, a través de métricas de bondad de ajuste y capacidad predictiva.
- Elaborar recomendaciones para la entidad de garantías respecto al uso del modelo en sus procesos de gestión y toma de decisiones.

4. Marco Teórico

4.1 Fundamentos del Riesgo de Crédito

El riesgo de crédito cuantifica la expectativa de pérdidas derivadas del incumplimiento de un deudor y se descompone, según Basilea III, en tres componentes principales: la probabilidad de incumplimiento (PD, Probability of Default), la exposición al incumplimiento (EAD, Exposure at Default) y la pérdida dada el incumplimiento (LGD, Loss Given Default). La pérdida esperada (EL, Expected Loss) se define como:

$$\mathrm{EL} = \mathrm{PD} \times \mathrm{EAD} \times \mathrm{LGD},$$

donde la tasa de recuperación (RR, Recovery Rate) se expresa como

$$\mathrm{RR} = 1 - \mathrm{LGD},$$

indicando la proporción de exposición recuperada tras un incumplimiento.

En la literatura financiera, múltiples estudios han aplicado modelos estadísticos avanzados para predecir tasas de recuperación. Por ejemplo, Bonini y Caivano (2015) emplearon modelos de LGD basados en regresión lineal y técnicas de scorecards para estimar la tasa de recuperación en portafolios minoristas de Latinoamérica, capturando heterogeneidad en extremos y comportamientos atípicos; Crosbie y Bohn (2003) formularon modelos extendidos de riesgo de incumplimiento basados en máxima verosimilitud penalizada; Engelmann, Hayden y Tasche (2003) aplicaron medidas de potencia discriminativa para mejorar la predicción de LGD en préstamos corporativos; y Ospina y Ferrari (2010) introdujeron la distribución Beta Inflada en Cero y Uno (BEINF) dentro del marco GAMLSS para modelar proporciones con acumulaciones

en 0 y 1, ajustando simultáneamente las probabilidades de extremos y la parte continua mediante enlaces logit y log.

Estos enfoques demuestran que, para estimar con precisión la recuperación de crédito, resulta esencial seleccionar distribuciones que reflejen la heterogeneidad del proceso de cobro y utilizar técnicas que permitan modelar simultáneamente la media, la varianza y las probabilidades de extremos, tal como lo permite el enfoque GAMLSS-BEINF.

4.2 Modelado de Variables Proporcionales

4.2.1 Distribución Beta Clásica

La distribución Beta es adecuada para modelar variables continuas acotadas en el intervalo $(0,1)$, ya que su forma puede adoptar patrones simétricos o asimétricos y permitir heterocedasticidad vinculada al valor esperado. Ferrari y Cribari-Neto (2004) formalizaron la regresión Beta al parametrizar la densidad en términos de la media μ y la precisión ϕ, lo que separa explícitamente el nivel y la dispersión del proceso estocástico. La función de densidad de $Y \sim \text{Beta}(\alpha, \beta)$ se expresa como

$$f(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, y^{\alpha-1}(1-y)^{\beta-1}, 0 < y < 1,$$

donde $\alpha, \beta > 0$ (Smithson & Verkuilen, 2006). Al parametrizar $\alpha = \mu\phi$ y $\beta = (1-\mu)\phi$, se obtiene:

$$E(Y) = \mu, \text{Var}(Y) = \frac{\mu(1-\mu)}{1+\phi},$$

y la regresión se especifica mediante un enlace $\text{logit}(\mu) = \text{x}^\top\beta$, mientras ϕ puede modelarse como $\log(\phi) = \text{z}^\top\gamma$ para capturar heterocedasticidad (Ferrari & Cribari-Neto, 2004; Smithson & Verkuilen, 2006).

4.2.2 Distribución Beta Inflada en Cero y Uno (BEINF)

La Beta Inflada en Cero y Uno (BEINF) extiende la regresión Beta clásica para variables proporcionales que presentan acumulaciones puntuales en los límites 0 o 1, incorporando explícitamente probabilidades de ceros o unos junto con una componente continua en (0,1) (Ospina & Ferrari, 2010; Ospina & Ferrari, 2012). La función de densidad mixta se define como

$$f(y) = \begin{cases} \pi_0, & y = 0, \\ \pi_1, & y = 1, \\ (1 - \pi_0 - \pi_1)\, f_{\text{Beta}}(y; \alpha, \beta), & 0 < y < 1, \end{cases}$$

donde $\pi_0 = \Pr(Y = 0)$, $\pi_1 = \Pr(Y = 1)$ y $f_{\text{Beta}}(\cdot)$ es la densidad Beta parametrizada en media μ y precisión ϕ mediante $\alpha = \mu\phi$ y $\beta = (1 - \mu)\phi$ (Ospina & Ferrari, 2010).

El modelo BEINF se especifica a través de enlaces separados para cada componente:

$\text{Logit}(\mu) = \text{x}^\top\beta$ para la media continua,

$\text{Log}(\phi) = \text{z}^\top\gamma$ para la precisión,

$\text{Logit}(\pi_0) = \text{w}_0^\top\delta_0$ y $\text{Logit}(\pi_1) = \text{w}_1^\top\delta_1$ para las probabilidades de cero y uno, respectivamente (Ospina & Ferrari, 2012; Rigby et al., 2019).

4.3 Marco GAMLSS para Regresión Distribucional

Los modelos lineales generalizados (GLM) presentan limitaciones al restringir la distribución de la variable respuesta a la familia exponencial, mientras que los modelos aditivos

generalizados (GAM) únicamente incorporan funciones suavizadoras para modelar la media. Para superar estas restricciones, Rigby y Stasinopoulos (2005) desarrollaron los modelos aditivos generalizados para localización, escala y forma (GAMLSS, por sus siglas en inglés: Generalized Additive Models for Location, Scale and Shape), los cuales permiten modelar simultáneamente todos los parámetros de distribución (localización, escala y forma) a partir de una amplia variedad de distribuciones, ofreciendo flexibilidad para abordar asimetría, curtosis y colas gruesas en los datos.

El marco GAMLSS facilita la modelación de todos los parámetros de una familia extensa de distribuciones como funciones aditivas de covariables (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007). Para cada parámetro θ_k (k = 1, 2, ..., K), el modelo especifica un predictor mediante una función de enlace monotónica g_k(·) que relaciona el parámetro con las variables explicativas (Rigby & Stasinopoulos, 2005):

$$g_k(\theta_{k,i}) = \eta_{k,i} = \mathrm{x}_{k,i}^{\top}\beta_k + \sum_j s_{kj}(z_{kj,i})$$

donde θ{k,i} representa el k-ésimo parámetro de distribución para la i-ésima observación, η{k,i} es el predictor lineal, x{k,i} es un vector de covariables lineales, βk es el vector de coeficientes correspondiente, y s{kj}(·) representa funciones suavizadoras no paramétricas aplicadas a las covariables z{kj,i} (Stasinopoulos & Rigby, 2007). Esta especificación permite capturar relaciones complejas tanto lineales como no lineales entre las variables explicativas y los distintos parámetros de localización, escala y forma de la distribución condicional de la variable respuesta (Rigby & Stasinopoulos, 2005).

Para la distribución BEINF (Beta Inflada), utilizada en el modelado de tasas de recuperación acotadas en el intervalo con inflación en los puntos extremos, los parámetros son μ (localización), σ (escala), ν (inflación en cero) y τ (inflación en uno) (Rigby et al., 2019). La función de enlace g_k(·) es específica para cada parámetro según su rango natural: logit(μ) para el parámetro de localización acotado en (0,1), log(σ) para el parámetro de escala positivo, y logit(ν) y logit(τ) para los parámetros de inflación acotados en (0,1) (Rigby et al., 2019).

Esta estructura flexible del marco GAMLSS permite incorporar efectos tanto paramétricos como no paramétricos en el modelado simultáneo de todos los parámetros distribucionales, superando las limitaciones de los modelos tradicionales que únicamente modelan la media condicional (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007).

4.4 Estimación y Diagnóstico en GAMLSS

4.4.1 Algoritmo de Estimación Rigby-Stasinopoulos (RS)

El algoritmo de Rigby–Stasinopoulos (RS) está diseñado para estimar de manera eficiente los parámetros de un modelo GAMLSS mediante el enfoque de verosimilitud penalizada combinado con un esquema iterativo de *backfitting* (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007). El objetivo es maximizar la función de verosimilitud penalizada

$$\ell_p = \ell(\boldsymbol{\theta}) - \frac{1}{2}\sum_{k=1}^{K} \lambda_k \, \boldsymbol{\beta}_k^{\top} \mathbf{S}_k \, \boldsymbol{\beta}_k,$$

Donde $\ell(\theta)$ es la verosimilitud logarítmica del modelo sin penalización, $\theta = (\theta_1, \ldots, \theta_K)$ agrupa los K parámetros distribucionales, λ_k es el parámetro de suavizado para el k-ésimo predictor, β_k

son los coeficientes del predictor y S_k la matriz de penalización asociada al término de suavización (Rigby & Stasinopoulos, 2005).

La implementación RS emplea un esquema de *backfitting* en tres niveles (Rigby & Stasinopoulos, 2005):

1. Actualización externa de parámetros de suavizado: Se optimizan los parámetros λ_k utilizando criterios basados en el criterio de información generalizado ajustado (GAIC) o similares, ajustando la complejidad del modelo para cada parámetro de distribución.
2. Ajuste interno de efectos lineales y no lineales: Con λ_k fijos, cada predictor $g_k(\theta_k) = \eta_k$ se actualiza secuencialmente mediante métodos de máxima verosimilitud penalizada, resolviendo internamente un GLM extendido para cada parámetro condicional a los demás (Stasinopoulos & Rigby, 2007).
3. Estimación de suavizadores: Dentro de cada paso de GLM, se ajustan las funciones suavizadoras $s_{kj}(\cdot)$ a través de penalización cuadrática de base de funciones, imponiendo restricciones de suavidad mediante las matrices S_k y ajustando los coeficientes β_k mediante descomposición de matrices de penalización (Rigby & Stasinopoulos, 2005).

Este proceso iterativo continúa hasta cumplir criterios de convergencia tanto en los valores de β_k como en los parámetros de suavizado λ_k. La ventaja de RS radica en su capacidad para descomponer la compleja optimización multivariada en subproblemas más sencillos, manteniendo el control de la complejidad del modelo a través de penalizaciones adaptativas en cada parámetro distribucional.

4.4.2 Selección de Variables y Modelos

La gran cantidad de términos potenciales en un modelo GAMLSS requiere un procedimiento de selección automatizado que equilibre ajuste y complejidad. La función stepGAIC implementa un algoritmo de *stepwise* basado en criterios de información —AIC (κ = 2), BIC (κ = ln n) o GAIC con valores intermedios— para añadir o eliminar términos de forma secuencial según la contribución marginal de cada predictor al ajuste global (Stasinopoulos & Rigby, 2007). En cada paso, el modelo candidato se evalúa mediante:

$$\mathrm{GAIC}(\kappa) = -2\ell + \kappa\ \mathrm{edf},$$

donde ℓ es la verosimilitud logarítmica del modelo y edf son los grados de libertad efectivos totales (Stasinopoulos & Rigby, 2007).

Para restringir la selección de distribuciones al rango (0,1) —por ejemplo, al trabajar con proporciones o tasas— la función modelRange emplea el argumento `type = "real0to1"`, de modo que solo se consideran familias de distribuciones con soporte en el intervalo unitario (Rigby et al., 2019). Este enfoque es especialmente útil cuando el análisis requiere modelar simultáneamente varios parámetros (localización μ, escala σ, inflación en 0 y 1) en el contexto de distribuciones Beta Inflada (BEINF), garantizando coherencia en los dominios de cada parámetro (Rigby & Stasinopoulos, 2005; Rigby et al., 2019).

En síntesis, la combinación de stepGAIC para la selección de términos explicativos y modelRange(type = "real0to1") para acotar las familias distribucionales permite un flujo de trabajo integrado que optimiza tanto la especificación de covariables como la elección de la distribución más adecuada dentro del marco GAMLSS (Stasinopoulos & Rigby, 2007; Rigby et al., 2019).

4.4.3 Diagnostico de Residuos

En GAMLSS, los métodos de diagnóstico clásicos basados en residuos deviance o Pearson no son adecuados cuando la distribución de la respuesta se aleja de la normalidad o de la familia exponencial. Para abordar esta limitación, Dunn y Smyth (1996) introdujeron los residuos cuantílicos aleatorios, también denominados *randomized quantile residuals*, que transforman la verosimilitud acumulada de cada observación en residuos que, bajo el modelo correcto, se distribuyen aproximadamente como una normal estándar. Definidos para la i-ésima observación como

$$r_i = \Phi^{-1}\{F(y_i; \hat{\theta})\},$$

estos residuos incorporan una aleatorización uniforme cuando la función de distribución F es discontinua, garantizando así la continuidad del residual y la validez de los diagnósticos basados en normalidad (Dunn & Smyth, 1996).

En el marco GAMLSS, donde cada parámetro distribucional (μ, σ, ν, τ, etc.) se modela simultáneamente, los residuos cuantílicos aleatorios permiten evaluar de forma conjunta la adecuación de la especificación de todos los parámetros. Un diagrama de cuantiles–cuantiles tradicional de estos residuos revela discrepancias globales frente a la normalidad, mientras que el worm plot —un gráfico de residuos estandarizados en función de percentiles agrupados— facilita la detección de desviaciones locales en asimetría y curtosis (van Buuren & Fredriks, 2001). En el *worm plot*, cada “gusano” representa el sesgo medio de los residuos en un segmento de la distribución, de modo que curvas ascendentes indican asimetría positiva, descendentes asimetrías negativas, y curvaturas en “U” o “∩” revelan exceso de curtosis o colas gruesas.

La implementación de estos diagnósticos en GAMLSS (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007) permite:

- Verificar la validez del supuesto de distribución condicional mediante la aproximación a $\mathcal{N}(0,1)$ de los residuos cuantílicos.
- Detectar estructuras locales de falta de ajuste en cada zona de la distribución a través del worm plot.
- Evaluar la calibración de los parámetros de forma (ν, τ) en distribuciones inflacionadas y colas asimétricas.

4.4.4 Criterios de Bondad de Ajuste

Para evaluar la calidad del ajuste en el marco GAMLSS, además de los criterios de información (AIC, BIC, GAIC) se emplean medidas que exploran todas las dimensiones de la distribución condicional, no sólo la media.

Primero, la deviance generalizada compara el ajuste del modelo con el de un modelo "saturado" que reproduce exactamente los datos:

$$\text{Deviance} = -2[\ell(\text{saturado}) - \ell(\hat{\theta})],$$

donde $\ell(\hat{\theta})$ es la verosimilitud lograda por el modelo estimado y $\ell(\text{saturado})$ la verosimilitud máxima posible (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007). Un valor de deviance cercano a cero indica un ajuste muy próximo al saturado.

Adicionalmente, se calculan scores de predicción probabilística, como el Continuous Ranked Probability Score (CRPS), que mide la discrepancia entre la función de distribución acumulada pronosticada F y la realización observada y:

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} \{F(z) - \mathbf{1}(z \geq y)\}^2 \, dz.$$

El CRPS penaliza errores de predicción en toda la distribución condicional y es especialmente útil en contextos donde la varianza y la forma juegan un papel crucial (Gneiting & Raftery, 2007).

En el caso de modelos BEINF (Beta Inflada), que combinan una distribución continua en (0,1) con masas en 0 y 1, la deviance se descompone en tres componentes:

Verosimilitud de la inflación en cero (ν)

Verosimilitud de la inflación en uno (τ)

Verosimilitud de la parte Beta para valores en (0,1)

Esto permite examinar por separado el ajuste en los extremos y en la parte continua (Rigby et al., 2019). Asimismo, el CRPS para BEINF se adapta incorporando la masa discreta en los puntos 0 y 1, garantizando que los errores de probabilidad en los extremos se penalicen adecuadamente (Grimit et al., 2006).

En conjunto, la deviance y el CRPS ofrecen una visión complementaria: la deviance evalúa el ajuste global relativo al modelo saturado, mientras que el CRPS cuantifica la precisión probabilística en toda la distribución, incluyendo colas y polos inflados. Estos criterios resultan fundamentales en aplicaciones de recuperación de crédito, donde la correcta modelación de extremos (tasa cero o plena recuperación) y la dispersión de las tasas intermedias son críticas para la toma de decisiones (Rigby & Stasinopoulos, 2005; Gneiting & Raftery, 2007; Rigby et al., 2019).

4.5 Ventajas sobre Metodologías Alternativas

A diferencia de los GLM, que restringen la respuesta a la familia exponencial, y de los GAM, que únicamente suavizan la media condicional, GAMLSS ofrece un marco de regresión distribucional con las siguientes ventajas clave:

- Modelación simultánea de todos los parámetros distribucionales Mientras GLM y GAM solo explican la media, GAMLSS permite especificar funciones aditivas para la localización (μ), la escala (σ) y los parámetros de forma (por ejemplo, asimetría ν y curtosis τ), capturando simultáneamente el comportamiento de la media, la variabilidad y las colas de la distribución condicional (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007).
- Flexibilidad de familias más allá de la exponencial GAMLSS incorpora cientos de distribuciones continuas y discretas, incluidas aquellas con colas pesadas y formas asimétricas que no se ajustan a la familia exponencial, ampliando el conjunto de problemas abordables en comparación con GLM/GAM tradicionales (Rigby & Stasinopoulos, 2005).
- Ajuste de inflaciones y colas extremas Mediante distribuciones con componentes discretos e inflacionados (p. ej., BEINF para datos en con masas en los extremos), GAMLSS modela explícitamente fenómenos de inflación en ceros o unos que no pueden capturar GLM/GAM estándar (Rigby et al., 2019).
- Selección integrada de suavizadores y familias Con procedimientos de selección paso a paso (stepGAIC) y restricciones en el rango de las distribuciones (type = "real0to1"), GAMLSS automatiza la especificación de términos lineales, no lineales y de la familia más adecuada para cada parámetro, optimizando ajuste y complejidad simultáneamente (Stasinopoulos & Rigby, 2007).

- Diagnóstico detallado de ajuste Además de criterios de información, GAMLSS emplea residuos cuantílicos aleatorios y gráficos como el worm plot para evaluar localmente asimetría y curtosis, ofreciendo diagnósticos más informativos que los residuos deviance o Pearson en modelos no gaussianos (Dunn & Smyth, 1996; van Buuren & Fredriks, 2001).

Estas capacidades convierten a GAMLSS en una herramienta poderosa para el análisis de tasas de recuperación, donde la correcta modelación de la dispersión, la asimetría y la inflación en los extremos resulta esencial para obtener pronósticos precisos y decisiones informadas.

5. Metodología

Esta investigación adopta un enfoque cuantitativo para desarrollar un modelo predictivo que mejore la recuperación de créditos siniestrados. El diseño correlacional permite identificar relaciones entre variables explicativas y la tasa de recuperación, mientras que el componente predictivo facilita la implementación práctica del modelo.

5.1 Población y Muestra

Población objetivo: Créditos de consumo siniestrados respaldados por garantías

Marco muestral: Base de datos del fondo de garantía

Tamaño de muestra: 11.031 operaciones siniestradas entre 2013 a 2023

Tabla 1

Descripción de las variables incluidas en el estudio

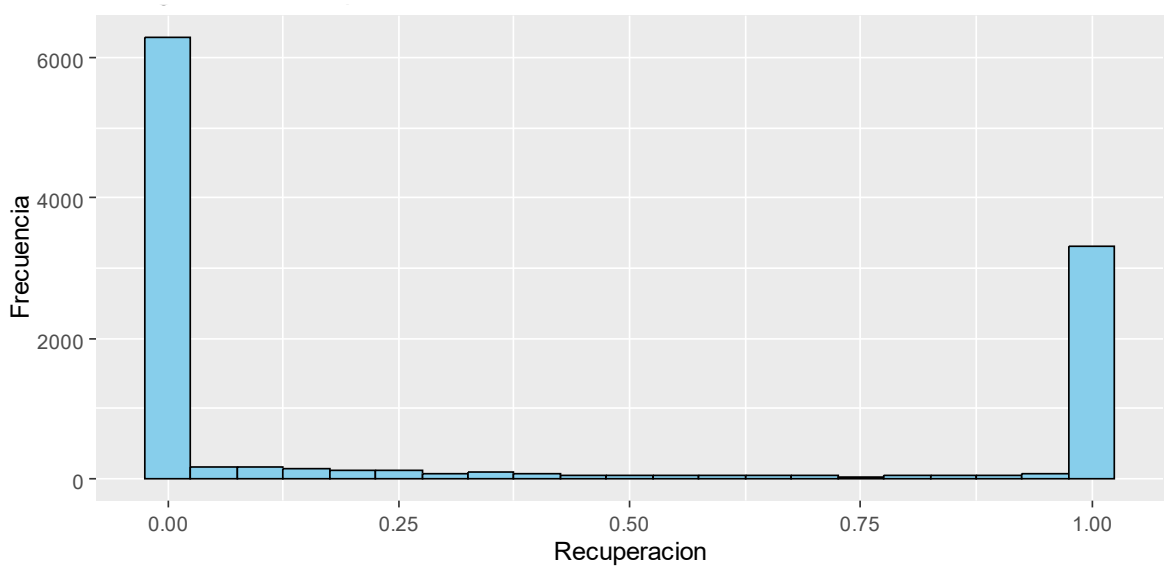
Nombre de la variable	descripción	Tipo de variable
Recuperación	Tasa de recuperación entre 0 y 1	Cuantitativa
Pago garantía	Valor pagado por la operación siniestrada	Cuantitativa
Plazo	Se refiere al período de tiempo acordado en meses para el cumplimiento total de una obligación financiera	Cuantitativa
proporción deuda	Expresa la relación porcentual entre el saldo pendiente de la deuda de un siniestro y el valor total original del crédito asociado.	Cuantitativa
Altura gestión	Representa la cantidad de meses transcurridos entre la fecha de ocurrencia del siniestro y la fecha del último corte de información.	Cuantitativa
Cantidad abonos	Registra el número total de pagos realizados al Fondo de Garantías después de la ocurrencia del siniestro.	Cuantitativa
Días mora	D: si la garantía se pagó en 90 días de mora E: si la garantía se pagó en 180 días de mora	Cualitativa
Recuperación temprana	sí: la operación tuvo un abono antes de 12 meses de gestión no: la operación no tuvo un abono antes de 12 meses de gestión	Cualitativa

5.2 Procedimiento de Análisis

5.2.1 Análisis Exploratorio de Datos

Figura 1.

Histograma tasa de recuperación



La variable Recuperación representa la proporción recaudada sobre el saldo adeudado al momento del siniestro ($0 \leq R \leq 1$). La distribución observada en la Figura 1 muestra inflación en los extremos: 55.9% casos con R=0, 14.3% con R=1, y 29,8% con recuperación parcial, justificando el uso de la distribución BEINF

El análisis descriptivo en la Tabla 2 de las variables explicativas revela características distribucionales relevantes que influyen en la especificación del modelo econométrico propuesto.

Tabla 2

Resultados descriptivos

Variables	mínimo	1° cuartil	2° cuartil	Promedio	3° cuartil	Máximo
Pago garantía (millones)	0,006	1,0	1,79	2,63	3,32	56,2
Plazo (meses)	6	18	24	31	36	120
Proporción deuda	0,0029	0,5848	0,7567	0,7302	0,9109	1

Altura gestión	25	46	73	73	98	162
Cantidad abonos	0	0	0	1	1	35

El monto pagado por el fondo de garantías presenta una distribución marcadamente asimétrica, con valores comprendidos entre $0,006 millones y $56,2 millones La diferencia entre la mediana ($1,79 millones) y la media ($2,63 millones) evidencia la presencia de observaciones extremas que sesgan la distribución hacia valores superiores. Esta asimetría positiva es característica de variables financieras en contextos de riesgo crediticio.

La variable Plazo, que refleja la duración contractual originalmente pactada, oscila entre 6 y 120 meses, con una mediana de 24 meses y una media de 31 meses. El 75% de los créditos se concentra en plazos de hasta 36 meses, indicando una predominancia de productos de corto y mediano plazo en la cartera.

La Proporción de Deuda, que mide la relación entre el saldo insoluto y el monto original del crédito, varía entre 0,0029 y 1,0000. Con mediana en 0,7567 y tercer cuartil en 0,9109, esta variable refleja que la mayoría de las operaciones presenta un alto nivel de endeudamiento relativo, lo cual puede influir en la probabilidad de recuperación.

Altura gestión, indicador temporal de la gestión de recuperación de la garantía, abarca valores entre 25 y 162 meses, con mediana de 73 meses y distribución relativamente uniforme. Esta variable permitirá evaluar efectos temporales en los procesos de recuperación.

Cantidad de Abonos, que contabiliza los pagos realizados tras el siniestro, varía entre 0 y 35, con mediana de 0 y media de 1. La alta concentración en cero evidencias que la mayoría de los

casos no registra abonos durante el periodo de análisis, destacando la importancia de incorporar mecanismos que capten las dinámicas de pago residual.

Recuperación Temprana, variable dicotómica, indica que 2969 casos presentan recuperación en fases iniciales y 8062 no la registran, lo que reafirma la heterogeneidad en los tiempos y modalidades de recuperación.

Categoría clasifica las operaciones en tipologías diferenciadas, con 6466 observaciones en la categoría D y 4565 en la categoría E. Este agrupamiento permite controlar la heterogeneidad no observada asociada al pago de la garantía.

Como parte del análisis exploratorio de datos, se implementó una matriz de correlación de Spearman para examinar las asociaciones bivariadas entre las variables explicativas y la variable dependiente recuperación. La selección del coeficiente de correlación de Spearman se fundamentó en las características distribucionales de los datos, que presentan asimetría y la presencia de valores atípicos, condiciones que hacen inadecuada la aplicación de medidas paramétricas como el coeficiente de Pearson.

Figura 2

Análisis de correlaciones

	pago garantía	plazo	proporción deuda	altura gestión	cantidad abonos	recuperación
pago garantía	1	0,61	0,48	-0,06	-0,15	-0,21
plazo	0,61	1	0,13	0,13	-0,17	-0,16
proporción deuda	0,48	0,13	1	-0,14	-0,2	-0,25
altura gestión	-0,06	0,13	-0,14	1	0,04	0,16
cantidad abonos	-0,15	-0,17	-0,2	0,04	1	0,91
recuperación	-0,21	-0,16	-0,25	0,16	0,91	1

Los resultados de correlaciones en la Figura 2 evidencian una asociación positiva muy fuerte entre la cantidad de abonos y la tasa de recuperación ($\rho = 0.91$), lo que sugiere que esta variable constituye un predictor fundamental del comportamiento de recuperación crediticia. Por el contrario, se observa una correlación negativa moderada entre la proporción de deuda y la recuperación ($\rho = -0.25$), indicando que mayores niveles de endeudamiento relativo se asocian con menores tasas de recuperación.

La variable pago garantía presenta una correlación negativa débil con la recuperación ($\rho = -0.21$), lo que indica que montos más elevados pagados por el fondo de garantías por concepto de siniestros se asocian con menores tasas de recuperación posterior. Similarmente, el plazo muestra una asociación negativa menor ($\rho = -0.16$), sugiriendo que créditos originados con plazos contractuales más extensos tienden a presentar menores tasas de recuperación. Esta relación podría explicarse por el hecho de que los créditos a largo plazo pueden estar asociados con mayor exposición al riesgo y deterioro de la capacidad de pago del deudor a lo largo del tiempo, lo que posteriormente impacta la efectividad de los procesos de recuperación.

La identificación de outliers constituye una etapa fundamental para caracterizar la dispersión y la heterogeneidad de los datos, especialmente en contextos financieros donde comportamientos inusuales en riesgo es habitual.

El procedimiento se desarrolló mediante el método intercuartílico (IQR), aplicado de forma automatizada sobre cada variable numérica independiente. Este enfoque emplea el criterio de los boxplots, clasificando como valores atípicos aquellas observaciones que se sitúan fuera del rango definido por $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$.

Sobre el conjunto de operaciones siniestradas, compuesto por 11.031 observaciones los resultados muestran una cantidad relevante de registros extremos en las variables "Pago Garantía" (751 outliers), "Plazo" (165 outliers), "Proporción_deuda" (26 outliers) y, de forma destacada, "Cantidad_abonos" (1.263 outliers). En contraste, "mes_corte_siniestro" y "Recuperación" no presentan registros atípicos bajo el criterio empleado. Este patrón es coherente con la naturaleza de los datos analizados, en tanto las operaciones de crédito siniestradas suelen registrar comportamientos altamente heterogéneos.

Cabe resaltar que, dado el tamaño de la muestra y el hecho de que todos los datos corresponden a operaciones reales, la presencia de valores atípicos no necesariamente implica errores o inconsistencias. Por el contrario, constituye un reflejo de la diversidad y complejidad inherente al fenómeno crediticio y a los procesos de recuperación asociados al riesgo de la operación. En consecuencia, se optó por conservar todos los registros en el análisis subsecuente, complementando las técnicas convencionales con métodos estadísticos robustos y transformaciones adecuadas cuando resulta pertinente.

5.2.2 Estructura General del Modelo GAMLSS-BEINF

Para una tasa de recuperación R_i de la i-ésima operación crediticia siniestrada, el modelo se especifica como:

$$R_i \sim BEINF(\mu_i, \sigma_i, \nu_i, \tau_i)$$

5.2.3 Selección del Modelo Distribucional

La selección de la distribución adecuada es un proceso central en la modelización estadística de variables proporcionales acotadas, como la tasa de recuperación observada en operaciones de cartera siniestrada. Para cumplir con este objetivo metodológico, se desarrolló una

estrategia exhaustiva de comparación de distribuciones candidatas utilizando la función chooseDist del paquete GAMLSS, garantizando así la robustez y validez del modelo final.

El procedimiento inició con el ajuste de un modelo base (modelo0) bajo la familia Beta Inflada (BEINF), considerando únicamente interceptos en todos los parámetros del modelo. Posteriormente, se empleó la rutina chooseDist para evaluar de manera sistemática un conjunto de distribuciones idóneas para variables definidas en el rango, utilizando el argumento type = "real0to1". Este filtro asegura que la búsqueda se circunscriba a familias distribucionales capaces de capturar tanto los valores intermedios como la posible inflación en los extremos de la variable respuesta.

La comparación de cada distribución se realizó aplicando el Generalized Akaike Information Criterion (GAIC) bajo distintos valores del parámetro de penalización (k = 2, 3.84, 9.31), equivalentes a los criterios AIC, test chi-cuadrado y BIC respectivamente. Esta multiplicidad de penalizaciones permite evaluar el balance entre ajuste y parsimony según diversos estándares de rigurosidad estadística.

Figura 3

Comparación de distribuciones

minimum GAIC(k= 2) family: BEINF
minimum GAIC(k= 3.84) family: BEINF
minimum GAIC(k= 9.31) family: BEINF

Type "real0to1"	2	3.84	9.31
BE	NA	NA	NA
BEo	NA	NA	NA
BEINF0	NA	NA	NA
BEINF1	NA	NA	NA
LOGITNO	NA	NA	NA
SIMPLEX	NA	NA	NA
BEOI	NA	NA	NA
BEZI	NA	NA	NA
BEINF	20734.58	20741.94	20763.82
GB1	NA	NA	NA

Los resultados en la Figura 3 demostraron consistentemente que la familia BEINF presenta los valores de GAIC más bajos en los tres escenarios evaluados. Este hallazgo revela la superioridad de la distribución Beta Inflada para modelar la variable de recuperación crediticia, ya que permite capturar efectivamente tanto la acumulación en los extremos (cero y uno) como la dispersión completa sobre el intervalo unitario. Es importante señalar que las demás familias distribucionales arrojaron valores NA en la evaluación, lo que indica que dichas distribuciones no lograron ajustarse a la naturaleza específica de los datos.

5.2.4 Selección de Variables

La especificación óptima de un modelo GAMLSS requiere identificar, para cada uno de sus componentes, el subconjunto de variables explicativas que maximiza la parsimonia sin sacrificar la calidad de ajuste. En este estudio, se adoptó un enfoque de selección *stepwise* basado

en el criterio de información de Akaike (AIC) para los cuatro parámetros de la distribución Inflada Beta (μ, σ, ν y τ). A continuación, se procedió de manera independiente para cada parámetro:

Parámetro μ: Se aplicó la selección *stepwise* bidireccional para identificar las covariables que aportan información significativa al pronóstico de la media de la recuperación. El algoritmo evalúa adiciones y eliminaciones de variables con base en la mejora del AIC, deteniéndose cuando ningún cambio adicional reduce el valor del criterio.

$$logit(\mu_i) = \beta_0^{(\mu)} + \beta_1^{(\mu)} cantidad\ abonos_i + \beta_2^{(\mu)} proporción\ deuda_i + \beta_3^{(\mu)} recuperación\ temprana_i + \beta_4^{(\mu)} altura\ gestión_i$$

Parámetro σ: Para modelar la heterocedasticidad intrínseca en los procesos de recuperación, la selección de variables siguió un procedimiento equivalente, explorando aquellas covariables que explican la variabilidad de los residuos. El uso del AIC garantiza un balance adecuado entre complejidad y ajuste.

$$log(\sigma_i) = \gamma_0^{(\sigma)} + \gamma_1^{(\sigma)} altura\ gestión_i + \gamma_2^{(\sigma)} recuperación\ temprana_i$$

Parámetro ν: Dado el elevado porcentaje de observaciones con recuperación nula, se seleccionaron explicativas específicas para capturar los factores que determinan la probabilidad de observaciones en cero. El paso *stepwise* permitió excluir variables redundantes, optimizando la precisión del componente de inflación.

$$log\ (\nu_i) = \delta_0^{(\nu)} + \delta_1^{(\nu)} cantidad\_abonos_i + \delta_2^{(\nu)} altura\ gestión_i + \delta_3^{(\nu)} recuperación\ temprana_i + \delta_4^{(\nu)} proporción\ deuda_i + \delta_5^{(\nu)} días\ mora_i$$

Parámetro τ: Análogamente, se ejecutó la selección independiente para modelar la probabilidad de recuperación completa, asegurando que solo permanezcan en el modelo aquellas variables que contribuyen significativamente a explicar la inflación en el extremo superior de la distribución.

$$log(\tau_i) = \lambda_0^{(\tau)} + \lambda_1^{(\tau)} cantidad\ abonos_i + \lambda_2^{(\tau)} altura\ gestión_i + \lambda_3^{(\tau)} recuperación\ temprana_i$$
$$+ \lambda_4^{(\tau)} proporción\ deuda_i + \lambda_5^{(\tau)} pago\ garantía_i + \lambda_6^{(\tau)} días\ mora_i$$

Este esquema modular de selección de variables un procedimiento de *stepwise* AIC aplicado de forma separada a μ, σ, ν y τ permite construir un modelo robusto (Tabla 3), equilibrado y ajustado a las particularidades de los datos, al mismo tiempo que se minimiza el riesgo de sobreajuste al considerar el número óptimo de parámetros para cada componente.

Tabla 3

Funciones de Enlace

Parámetro	Variables Principales	Justificación Teórica
μ (media)	Cantidad abonos Proporción deuda Recuperación temprana Altura gestión	Determinan la capacidad típica de recuperación para casos con 0 < R < 1
σ (dispersión)	Altura gestión Recuperación temprana	Relacionan con la variabilidad del proceso de cobro
ν (inflación en 0)	Cantidad abonos Altura gestión Recuperación temprana Proporción deuda Días mora	Indican la probabilidad de pérdida total (R=0)
τ (inflación en 1)	Cantidad abonos Altura gestión Recuperación temprana	Señalan la probabilidad de recuperación completa (R=1)

	Proporción deuda
	Pago garantía
	Días mora

5.2.5 Estimación de parámetros y significancia estadística

Tabla 4

Estimaciones de parámetros, modelo BEINF

Parametro	μ		σ		ν		τ	
Link	logit		log		log		log	
	$\hat{\beta}^{\mu}$	SE ($\hat{\beta}^{\mu}$)	$\hat{\gamma}^{\sigma}$	SE ($\hat{\gamma}^{\sigma}$)	$\hat{\delta}^{\nu}$	SE ($\hat{\delta}^{\nu}$)	$\hat{\lambda}^{\tau}$	SE ($\hat{\lambda}^{\tau}$)
(Intercepto)	0,474**	0,159	-0,223**	0,081	8,110***	0,469	1.446***	0,172
Pago garantía							−0.000***	0,000
Plazo	-0,015***	0,001						
Proporción deuda	-1,216***	0,146			0,587	0,389	-1,376***	0,165
Altura gestión	0,003***	0,001	0,009***	0,000	-0.015***	0,002	0.009***	0,001
Cantidad abonos	0,173***	0,007			-8,219***	0,244	0.066***	0,009
Días mora_E					-0,645***	0,193	-0,2757***	0,078
Recuperación temprana_SI	-0,519***	0,069	-0,192***	0,055	-1,397***	0,183	−0.892***	0,070

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No. of observations in the fit:	11031
Degrees of Freedom for the fit:	22
Residual Deg. of Freedom:	11009
at cycle:	6
Global Deviance:	6042,902
AIC:	6086,902
SBC:	6247,688

En la Tabla 4 se detallan las estimaciones de los parámetros del modelo BEINF, junto a los respectivos errores estándar y niveles de significancia estadística, representados mediante asteriscos conforme a las convenciones usuales en análisis inferencial. Un mayor número de asteriscos refleja una menor probabilidad de que el valor estimado sea atribuible al azar, indicando una asociación estadísticamente robusta con la variable respuesta.

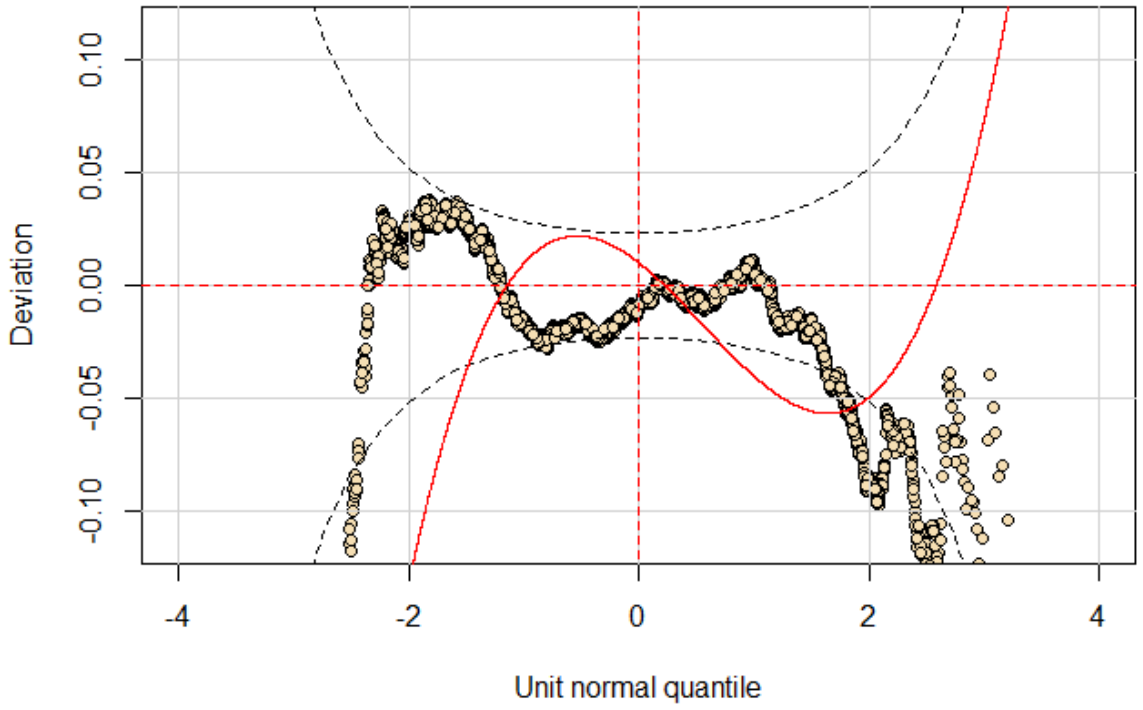
El análisis evidencia que varias covariables presentan estimaciones significativas en distintos parámetros del modelo (μ, σ, ν, τ). Por ejemplo, la proporción de deuda muestra un coeficiente negativo y altamente significativo en el parámetro de la media (μ), lo que sugiere una disminución esperada en la variable de interés ante un aumento en dicha proporción. De igual manera, altura de gestión y cantidad de abonos presentan efectos significativos sobre los parámetros de localización y dispersión, respectivamente. En contraste, las variables cuyos coeficientes carecen de significancia no muestran evidencia de asociación sistemática, bajo los niveles convencionales ($p < 0,05$).

En suma, la significancia estadística identificada respalda la inclusión de ciertos predictores en la especificación final del modelo, facilitando una interpretación precisa y fundamentada de los determinantes principales del proceso de recuperación cuantificado.

5.2.6 Diagnósticos de Bondad de Ajuste

Figura 4.

Worm plot modelo BEINF - tasa recuperación



La Figura 4 muestra una clara desviación en los extremos (colas) y ondulaciones en la zona central, con grupos de residuos alejados de la banda de tolerancia, especialmente en los cuantiles negativos y positivos más extremos.

Esto sugiere que el modelo BEINF, aunque puede capturar parte del comportamiento central de los datos, tiene dificultades para ajustar correctamente los valores extremos y posiblemente subestima o sobrestima la dispersión en ciertas regiones del espacio de predicción.

Es posible que las variables explicativas incluidas en el modelo no sean suficientes para captar toda la variabilidad, especialmente en los extremos; podría requerirse la inclusión de variables

adicionales, una especificación alternativa de la fórmula, o el uso de otro modelo de inflación para mejorar el ajuste.

6. Resultados

Los resultados del modelo BEINF muestran que la recuperación de créditos no depende de un solo proceso, sino de tres mecanismos distintos, cada uno con predictores y efectos específicos.

El modelo identificó dinámicas diferentes para cada componente de la recuperación, con patrones de significancia y magnitud que reflejan mecanismos diversos que influyen en el resultado final.

6.1 Componente μ: Recuperación Parcial

El componente μ, que modela la recuperación esperada en operaciones con resultados intermedios ($0 < R < 1$), estuvo determinado principalmente por dos variables predictoras: la cantidad de abonos y la proporción de deuda.

Cantidad de Abonos ($\beta = 0.173$, $p < 0,05$): La cantidad de abonos emergió como el predictor más favorable en este componente. Este resultado indica que cada abono adicional realizado post-siniestro incrementa las razones de probabilidades de recuperación en 18.9%.

Proporción de Deuda ($\beta = -1.216$, $p < 0,05$): Opuestamente, la proporción de deuda ejerció un efecto negativo sustancial, representando la segunda magnitud más importante en el componente μ. Cada unidad adicional de deuda pendiente reduce las razones de probabilidades en 70.4%. Este hallazgo subraya que el comportamiento del deudor antes del siniestro actúa como

una restricción estructural a la capacidad de recuperación, independientemente de los esfuerzos de gestión posteriores.

Plazo ($\beta = -0.015$, $p < 0,05$): El plazo demostró un efecto significativo aunque de menor magnitud, reduciendo la recuperación esperada aproximadamente en 0.4 puntos porcentuales por mes adicional.

Altura de Gestión ($\beta = 0.003$, $p < 0,05$): La altura de gestión mostró un efecto positivo marginal, indicando que la persistencia temporal de los esfuerzos de recuperación contribuye modestamente a mejores resultados.

Recuperación Temprana (Categoría SI, $\beta = -0.519$, $p < 0,05$): La recuperación temprana exhibió un efecto negativo paradójico, reduciendo la recuperación esperada en promedio. Este resultado contraintuitivo sugiere que operaciones donde se realizan pagos dentro del primer año post-siniestro puede representar casos de "satisfacción parcial" que no conducen a recuperación continuada.

6.2 Componente ν: Riesgo de Pérdida Total

El componente ν, modelando la probabilidad de no recuperar nada ($R = 0$), presentó un patrón radicalmente diferente, con cantidad de abonos como predictor absolutamente dominante.

Cantidad de Abonos ($\beta = -8.219$, $p < 0,05$): El coeficiente de -8.219 fue aproximadamente 10 veces mayor en magnitud que cualquier otro predictor en el modelo completo, evidenciando que este factor opera como un cambio cualitativo decisivo en la trayectoria de recuperación.

Esta gran magnitud implica que cada abono realizado reduce las razones de probabilidades de pérdida total en 99.973%, traduciéndose a cambios extremos en la probabilidad. En términos operacionales, esto significa que el primer abono es el punto de quiebre crítico en la trayectoria de

recuperación, separando casi completamente los casos de pérdida total de aquellos con alguna recuperación.

Recuperación Temprana (Categoría SI, $\beta = -1.397$, $p < 0,05$): La variable de recuperación temprana mostró un efecto negativo significativo, indicando que las operaciones que realizan pagos dentro del primer año post-siniestro presentan una reducción del 75.3% en el riesgo de pérdida total en comparación con aquellas que no realizan pagos tempranos. Este hallazgo resalta la importancia de los pagos iniciales como un determinante clave en el destino de la recuperación, sugiriendo que la recuperación temprana se asocia con un menor riesgo relativo.

Días de Mora (Categoría E, $\beta = -0.645$, $p < 0,05$): Las operaciones siniestradas a la categoría E, presentaron una razón de odds de pérdida total significativamente menor (un 47.6% de reducción) en comparación con las operaciones siniestradas (categoría D), según los resultados del modelo. Este hallazgo sugiere que, bajo las condiciones y el ajuste del modelo, el riesgo relativo de pérdida total es menor para las operaciones en mora más prolongada, aunque este resultado debe interpretarse considerando posibles efectos de control y otras variables incluidas en el análisis.

6.3 Componente τ: Probabilidad de Recuperación Total

El componente τ, modelando la probabilidad de recuperación completa ($R = 1$), reveló que la proporción de deuda es el factor estructural determinante.

Proporción de Deuda ($\beta = -1.376$, $p < 0,05$): El coeficiente negativo y de mayor magnitud respecto al componente μ indica que la proporción de deuda tiene un impacto especialmente fuerte en la probabilidad de recuperación total. Específicamente, un incremento completo en la proporción de deuda (de 0.1% a 100%) se asocia con una reducción del 74.74% en la probabilidad

de recuperación. Este hallazgo refleja una relación prácticamente estructural entre el hábito de pago previo al siniestro y la capacidad de recuperación completa, operando de manera independiente a otros factores incluidos en el modelo.

La recuperación temprana (Categoría "Sí", $\beta = -0.892$, $p < 0,05$) mostró un efecto negativo significativo, reduciendo en un 59% la probabilidad de recuperación total en comparación con las operaciones sin recuperación temprana. Este resultado sugiere que los pagos efectuados en el primer año suelen reflejar recuperaciones parciales más que una recuperación completa y definitiva.

Los días de Mora (Categoría E, $\beta = -0.276$, $p < 0,05$): Las operaciones clasificadas en la categoría E, con pagos posteriores a 180 días, presentan una reducción del 24.10% en la probabilidad de recuperación total en comparación con aquellas pagadas dentro de los 90 días (categoría D). Este resultado indica que un mayor tiempo en mora está asociado con una menor probabilidad de recuperar el crédito en su totalidad.

Cantidad de Abonos ($\beta = 0.066$, $p < 0,05$): Finalmente, cantidad de abonos mostró un efecto positivo, pero sorprendentemente débil, contrastando radicalmente con su coeficiente masivo en ν (-8.219). Esto sugiere una dinámica importante: mientras que los abonos son críticos para prevenir pérdida total, son casi irrelevantes para determinar recuperación total.

7. Conclusiones

Este estudio examinó los determinantes de la recuperación de créditos de consumo respaldados por garantías mediante un modelo de regresión Beta Inflada en Ceros y Unos (BEINF), utilizando datos de operaciones siniestradas de un Fondo de Garantías. El enfoque distribucional permitió descomponer el proceso de recuperación en tres componentes interdependientes.

Los resultados revelan que la recuperación de las operaciones siniestradas es un fenómeno multidimensional con dinámicas específicas para cada componente. En primer lugar, respecto al riesgo de pérdida total (componente ν), la cantidad de abonos post-siniestro emerge como predictora dominante con un coeficiente de magnitud (-8.219). Este hallazgo indica que operaciones sin pagos iniciales presentan una probabilidad de 99.97% de pérdida total, mientras que operaciones con tres abonos alcanzan prácticamente 0% de riesgo de pérdida. Adicionalmente, las operaciones con categoría E reduce el riesgo de pérdida total en un 47.53% comparada con la categoría D, sugiriendo que mora prolongada puede estar vinculada a negociaciones más sólidas.

En segundo término, en la recuperación parcial (componente μ), se identificaron dos predictores principales de magnitud opuesta: los abonos incrementan la recuperación en 18.9% por unidad ($\beta = 0.173$), mientras que la proporción de deuda previa la reduce significativamente en un 70.4% ($\beta = -1.216$). Este contraste evidencia que el comportamiento de pago del deudor antes del siniestro actúa como una restricción estructural incluso cuando hay esfuerzos de pago.

En tercer lugar, la probabilidad de recuperación total (componente τ) se determina casi exclusivamente por la proporción de deuda previa, con la recuperación temprana mostrando un efecto paradójico al reducir la probabilidad de recuperación total en un 59%. La categoría E también reduce esta probabilidad en un 24.10% respecto a la categoría D, indicando que la mora prolongada limita las perspectivas de recuperación completa.

Para investigaciones posteriores, es fundamental incorporar variables relacionadas con características del crédito inicial, tales como el monto originado, la tasa de interés y edad del deudor. Asimismo, resulta relevante incluir características contextuales por departamentos y zonas geográficas (urbanas y rurales) que permitan capturar heterogeneidad regional en los procesos de recuperación.

En síntesis, este trabajo demuestra que modelos distribucionales flexibles como BEINF constituyen herramientas poderosas para capturar la complejidad multidimensional de la recuperación crediticia, con implicaciones operacionales directas para la estrategia de gestión de cartera.

Referencias Bibliográficas

Asociación Europea de Fondos de Garantía (AECM). (2024). *AECM Annual Activity Report 2024.* AECM.

Asociación Latinoamericana de Instituciones de Desarrollo (ALIDE). (2018). *Estudio: Sistemas de garantías en América Latina.* ALIDE.

Basel Committee on Banking Supervision. (2017). *Basel III: Finalising post-crisis reforms.* Banco de Pagos Internacionales.

Bonini, C. P., & Caivano, M. (2015). *LGD forecasting and credit recovery: Evidence from Latin American markets. Journal of Credit Risk*, 11(2), 1–25.

Comisión Económica para América Latina y el Caribe (CEPAL). (2021). *Hacia un sistema nacional de garantías: antecedentes, mejores prácticas e implicancias para el caso argentino.* Naciones Unidas.

Consejo de Administración del Fondo de Garantías (CASFOG). (2024). *Informe del sistema de garantías (Abril 2024).* CASFOG.

Crosbie, P., & Bohn, J. (2003). *Modeling default risk.* KMV Corporation (White Paper).

Dunn, P. K., & Smyth, G. K. (1996). *Randomized quantile residuals. Journal of Computational and Graphical Statistics*, 5(3), 236–244. https://doi.org/10.1080/10618600.1996.10474708

Engelmann, B., Hayden, E., & Tasche, D. (2003). *Testing rating accuracy. Risk*, 16(2), 108–113.

Ferrari, S. L. P., & Cribari-Neto, F. (2004). *Beta regression for modelling rates and proportions. Journal of Applied Statistics*, 31(7), 799–815. https://doi.org/10.1080/0266476042000214501

Galvis, M. A. (2019). *Modelo de cálculo de la probabilidad de recuperación para entidades del sector cooperativo* [Tesis de maestría]. Universidad Autónoma de Bucaramanga.

Gneiting, T., & Raftery, A. E. (2007). *Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association*, 102(477), 359–378. https://doi.org/10.1198/016214506000001437

Grimit, E. P., Gneiting, T., Berrocal, V. J., & Johnson, N. A. (2006). *The continuous ranked probability score for circular variables and its application to mesoscale forecast verification. Quarterly Journal of the Royal Meteorological Society*, 132(620), 2925–2942. https://doi.org/10.1256/qj.05.195

Hossain, A., Rigby, R. A., Stasinopoulos, D. M., & Enea, M. (2016). *Centile estimation for a proportion response variable. Statistics in Medicine*, 35(6), 895–904.

Nacional Financiera (NAFIN). (2022). *Esquema de garantías NAFIN* [PDF]. NAFIN.

Ospina, R., & Ferrari, S. L. P. (2010). *Inflated beta distributions. Statistical Papers*, 51(1), 111–126. https://doi.org/10.1007/s00362-008-0133-y

Ospina, R., & Ferrari, S. L. P. (2012). *A general class of zero-or-one inflated beta regression models. Computational Statistics & Data Analysis*, 56(6), 1609–1623. https://doi.org/10.1016/j.csda.2011.11.008

Rigby, R. A., & Stasinopoulos, D. M. (2005). *Generalized additive models for location, scale and shape. Applied Statistics*, 54(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Rigby, R. A., Stasinopoulos, D. M., Heller, G. Z., & De Bastiani, F. (2019). *Distributions for modeling location, scale, and shape: Using GAMLSS in R.* CRC Press. https://doi.org/10.1201/9780429298547

Smithson, M., & Verkuilen, J. (2006). *A better lemon squeezer? Maximum-likelihood regression with Beta-distributed dependent variables. Psychological Methods*, 11(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Stasinopoulos, D. M., & Rigby, R. A. (2007). *Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software*, 23(7), 1–46. https://doi.org/10.18637/jss.v023.i07

Stasinopoulos, M., Enea, M., Rigby, R. A., & Hossain, A. (2017). *Inflated distributions on the interval. gamlss.inf* Package Documentation.

Stasinopoulos, M. D., Kneib, T., Klein, N., Mayr, A., & Heller, G. Z. (2024). *Generalized additive models for location, scale and shape: A distributional regression approach, with applications.* Cambridge University Press.

Superintendencia de Economía Solidaria. (2019). *Guía para el cálculo de los deterioros por riesgo de crédito.* Superintendencia de Economía Solidaria.

van Buuren, S., & Fredriks, M. (2001). *Worm plot: A simple diagnostic device for modelling growth reference curves. Statistics in Medicine*, 20(8), 1259–1277. https://doi.org/10.1002/sim.790