

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



**MODELO DE REGRESIÓN A LA MEDIA
SIMPLEX INFLACIONADA PARA PROPORCIONES**

**TESIS PARA OPTAR POR EL GRADO
DE MAGISTER EN ESTADÍSTICA**

Presentado por:

Juan Carlos Champác Flores

Asesor:

Dr. Cristian Luis Bayes Rodríguez

Miembros del jurado:

Dr. Luis Hilmar Valvidieso Serrano

Dr. Cristian Luis Bayes Rodríguez

Dr. Luis Enrique Benites Sánchez

Agosto, 2018

Dedicatoria

A mis padres Eleuterio y Gloria, incondicionales antes todo, siempre brindándome su invaluable apoyo y amor, que es la fuerza que me ayuda a seguir adelante cada día.

A mis hermanos Joe y Eric, por su comprensión, confianza y sobre todo, su ejemplo de superación.

A Zully, a quien admiro y amo cada día más por su gran fortaleza y humildad.



Agradecimientos

A mi asesor Cristian Bayes, por su constante apoyo y orientación brindada para la realización de esta tesis.

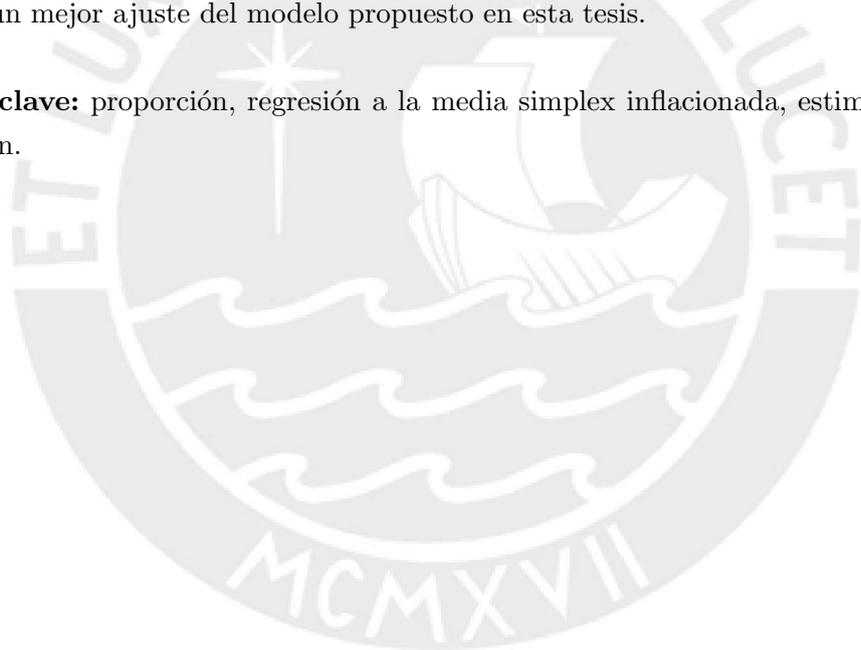
A Luis Valdivieso, José Flores, Enver Tarazona y Giancarlo Sal y Rosas, profesores de la Maestría de Estadística de la Pontificia Universidad Católica del Perú, quienes me han brindado su apoyo, conocimientos y orientación en los distintos cursos llevados durante la maestría.



Resumen

El presente trabajo de tesis propone el modelo de regresión a la media simplex inflacionada, que permite modelar variables aleatorias continuas limitadas en el intervalo cerrado $[0, 1]$ al considerar un conjunto de ecuaciones de regresión para estimar la media de la respuesta y los parámetros que modelan las probabilidades de los valores extremos 0 y 1. Asimismo, se desarrolla un estudio de simulación con el fin de evaluar si el método propuesto permite recuperar los parámetros del modelo desde el punto de vista de la estadística clásica. Por otro lado, se desarrolla la aplicación del modelo para determinar el grado de dolarización de empresas que registran deudas en el Sistema Financiero, y para evaluar el desempeño del mismo, se compara contra el modelo de regresión a la media beta inflacionada. Los resultados muestran un mejor ajuste del modelo propuesto en esta tesis.

Palabras-clave: proporción, regresión a la media simplex inflacionada, estimación clásica, dolarización.



Índice general

Índice de figuras	VII
Índice de cuadros	VIII
1. Introducción	1
1.1. Consideraciones Preliminares	1
1.2. Objetivos	8
1.3. Organización del Trabajo	8
2. Distribución Simplex Inflacionada	10
2.1. Distribución Simplex	10
2.1.1. Función de densidad de probabilidad	10
2.1.2. Propiedades	11
2.2. Distribución Simplex Inflacionada	12
2.2.1. Función de masa de probabilidad	12
2.2.2. Propiedades	13
2.3. Distribución Simplex Inflacionada Reparametrizada	13
2.3.1. Función de masa de probabilidad	13
2.3.2. Propiedades	15
3. Modelo de Regresión a la Media Simplex Inflacionada	16
3.1. Especificación del modelo de regresión a la media simplex inflacionada	16
3.1.1. Función de verosimilitud	17
3.1.2. Estimador de máxima verosimilitud	17
4. Estudio de Simulación	19
4.1. Consideraciones para la simulación	19
4.2. Resultados	21
5. Aplicación	23
5.1. El riesgo cambiario crediticio	23
5.2. Dolarización por cartera de créditos	24
5.3. Base de datos	24
5.4. Variables e indicadores	25
5.4.1. Información derivada de los estados financieros	25
5.4.2. Información derivada del Reporte Crediticio Consolidado	26

<i>ÍNDICE GENERAL</i>	VI
5.5. Resultados	27
6. Conclusiones	29
6.1. Conclusiones	29
6.2. Sugerencias para investigaciones futuras	29
A. Resultados teóricos	30
Bibliografía	34



Índice de figuras

2.1. Función de densidad de probabilidad de la distribución simplex con media μ (de izquierda a derecha: 0.3,0.5,0.7) y parámetro de dispersión σ^2 (de arriba a abajo: 49,25,5.33,0.25)	11
2.2. Relación de la varianza $V(W)$ contra el parámetro de dispersión σ^2 (gráfica de la izquierda, para distintos valores de $\mu = \{0.1, 0.3, 0.5, 0.6, 0.8\}$), y contra la media μ (gráfica de la derecha, para distintos valores de $\sigma^2 = \{0.25, 1, 4, 25, 100\}$), para una variable aleatoria que sigue una distribución simplex.	12
2.3. Función de densidad de probabilidad de la distribución simplex inflacionada reparametrizada con media γ (de izquierda a derecha: 0.3,0.3,0.8,0.8), parámetro de dispersión σ^2 (de izquierda a derecha: 1,6.25,1,6.25), y parámetros α_0 (de arriba a abajo: 0.3,0.3,0.8,0.8) y α_1 (de arriba a abajo: 0.3,0.8,0.3,0.8). . .	14
2.4. Relación de la varianza $V(Y)$ contra el parámetro de dispersión σ^2 (gráfica de la izquierda, para distintos valores de $\gamma = \{0.1, 0.3, 0.5, 0.6, 0.8\}$), y contra la media γ (gráfica de la derecha, para distintos valores de $\sigma^2 = \{0.25, 1, 4, 25, 100\}$), para una variable aleatoria que sigue una distribución simplex inflacionada reparametrizada.	15
4.1. Histograma de los datos, para $n = 2000$ valores simulados de y , bajo tres escenarios distintos: a) mayor proporción de valores 1 (gráfico superior izquierdo), b) mayor proporción de valores 0 (gráfico superior derecho) y c) similar proporción de valores 0 y 1 (gráfico inferior).	20
5.1. Histograma de la dolarización de créditos de deudores no minoristas	25

Índice de cuadros

4.1. Desempeño de los estimadores sesgo relativo, error cuadrático medio (ECM) y cobertura, bajo tres escenarios: 1) mayor proporción de valores unos, 2) mayor proporción de valores ceros y 3) similar proporción de valores ceros y unos, para $n = 500, 1000$ y 2000 valores simulados de y que siguen una distribución simplex inflacionada reparametrizada.	22
5.1. Coeficientes estimados para los modelos de regresión a la media simplex y beta inflacionadas. En paréntesis se muestra el error estándar de la estimación, obtenida de la inversa de la matriz de información de Fisher Observada evaluada en los parámetros estimados, y que permiten testear la significancia individual de los parámetros utilizando el estadístico de Wald (coeficientes significativos al 1% son señalados por **).	27
5.2. Criterios de comparación AIC, AICc, BIC, MSE y R^2 para los modelos de regresión a la media simplex y beta inflacionadas para el ajuste de los datos de dolarización.	28

Capítulo 1

Introducción

1.1. Consideraciones Preliminares

Es común considerar el modelo de regresión lineal para modelar una variable de interés en función de un conjunto de covariables. Esta relación se representa de la siguiente manera:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (1.1)$$

donde y_i es la variable dependiente, $\mathbf{x}_i^\top = [1, x_{i1}, x_{i2}, \dots, x_{ik}]^\top$ son las k variables independientes, $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]^\top$ son los k parámetros de regresión, ϵ_i es el término de error o perturbación y el subíndice i indica las observaciones muestrales por individuo. De esta manera, el objetivo del análisis no solo se basa en estimar los parámetros, sino de realizar inferencia estadística a fin de obtener resultados confiables a partir de un conjunto de observaciones. El estimador de mínimos cuadrados ordinarios (MCO) es el más utilizado, pero implica el cumplimiento de una serie de supuestos. En particular, se asume que $\epsilon_i \sim N(0, \sigma^2)$, es decir, proviene de una distribución normal con media cero y varianza constante (error homocedástico). Este supuesto, implica a su vez que la variable dependiente también se distribuye normalmente, es decir $y_i \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$.

Sin embargo, cuando se busca modelar una variable definida en el intervalo abierto $(0, 1)$ como una proporción o probabilidad, el estimador MCO deja de ser el más adecuado. Al respecto, [Kieschnick y McCullough \(2003\)](#) señalan que una proporción no se distribuye normalmente porque no se encuentra definida en toda la recta real \mathbb{R} , que es el dominio de la distribución normal. Además, señalan que la esperanza condicional debe ser una función no lineal de los parámetros, para que pueda estar definida en el intervalo limitado de la variable dependiente; y que la varianza condicional debe ser heterocedástica, al depender de la esperanza condicional (la varianza cambiará a medida que la esperanza se aproxime a los límites). En este sentido, una incorrecta especificación de la media y varianza condicional de la variable dependiente origina estimadores insesgados e inconsistentes para los coeficientes y errores estándares, por lo que son necesarias otras técnicas de estimación.

Una solución es transformar la variable y_i para que asuma valores en \mathbb{R} , y así modelar la media de la variable transformada mediante una regresión lineal. La transformación logit es la más usual y el planteamiento del modelo bajo este esquema es como sigue:

$$\ln\left(\frac{y_i}{1 - y_i}\right) = \tilde{y}_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad y_i \in (0, 1), \quad (1.2)$$

donde $\epsilon_i \sim N(0, \sigma^2)$. Sin embargo, tal como se señala en [Cribari-Neto y Zeileis \(2010\)](#), este enfoque presenta los siguientes problemas: a) la interpretación de los parámetros ya no es directa porque el enlace logit no se conserva para el valor esperado de la respuesta, es decir, los parámetros se interpretan en términos de la media de \tilde{y}_i y no de la media de y_i , b) la variable transformada, que proviene de proporciones, muestra generalmente heterocedasticidad al presentar mayor variación alrededor de la media y menor variación a medida que se acerca a los límites inferior y superior del intervalo, y c) las distribuciones de proporciones son típicamente asimétricas y, por tanto, las aproximaciones basadas en los supuestos de normalidad requeridos para la estimación de intervalos y pruebas de hipótesis pueden ser bastantes inexactas en muestras pequeñas.

Otra solución es considerar el modelo de regresión censurado, o modelo Tobit, donde la variable dependiente se encuentra censurada por encima o por debajo de algún valor, es decir, no se observa para una parte de la población. Para tal fin, se define una variable latente y_i^* , la cual se relaciona con la variable observada y_i mediante la siguiente relación:

$$y_i = \begin{cases} 0 & , y_i^* \leq 0, \\ y_i^* & , 0 < y_i^* < 1, \\ 1 & , y_i^* \geq 1. \end{cases}$$

Así, el modelo de regresión se define en términos de la latente, y se expresa como sigue:

$$y_i^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (1.3)$$

donde $\epsilon_i \sim N(0, \sigma^2)$, lo cual satisface los supuestos del modelo de regresión lineal. Sin embargo, es un modelo conceptualmente erróneo para modelar proporciones porque no se realiza una censura efectiva de los datos al no estar definido fuera del intervalo (0,1). Además, [Kieschnick y McCullough \(2003\)](#) señala que el modelo Tobit aplicado a proporciones es equivalente a un modelo de regresión normal y por tanto presenta las mismas deficiencias.

Otra alternativa más utilizada para el ajuste de estos datos es la distribución beta, cuya función de densidad es definida por:

$$b(y | a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, \quad y \in (0, 1),$$

donde $a > 0$, $b > 0$, Γ es la función gamma y con media igual a $E(y) = a/(a+b)$. Esta distribución es muy flexible para modelar proporciones debido a que la función de densidad de probabilidad puede tomar diversas formas dependiendo del valor de los parámetros a y b que la caracterizan, superando los problemas de heterocedasticidad y asimetría en los datos. Sin embargo, [Ferrari y Cribari-Neto \(2004\)](#) realizaron una reparametrización de la distribución considerando $\mu = a/(a+b)$ y $\phi = a+b$ con el fin de obtener una estructura de regresión simple para la media de la respuesta y que incluya un parámetro de dispersión, la cual será denota por $Y \sim Br(\mu, \phi)$ y cuya función de densidad viene dada por:

$$b(y | \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad y \in (0, 1), \quad (1.4)$$

donde $\mu \in (0, 1)$ es la media de la variable respuesta y $\phi > 0$ puede ser interpretado como un parámetro de dispersión debido a que $Var(y) = (\mu(1 - \mu))/(1 + \phi)$; existe una mayor dispersión de la distribución cuando ϕ disminuye, manteniendo constante la media. Asimismo, como señalan los autores, la distribución es simétrica cuando $\mu = 1/2$ y asimétrica cuando $\mu \neq 1/2$. En este sentido, si la variable y_i sigue esta última distribución, el modelo de regresión beta se puede representar de la siguiente manera:

$$\begin{aligned} Y_i &\sim Br(\mu_i, \phi), \\ g(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \mu_i \in (0, 1), \end{aligned} \tag{1.5}$$

donde μ_i es la media de la respuesta; $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ik}]^\top$ son vectores columna de covariables por individuo i ; $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_k]^\top$ son vectores columna de parámetros, y g es una función de enlace estrictamente monótona y diferenciable de segundo orden de $(0, 1)$ a \mathbb{R} . Este enfoque permite una interpretación más clara al modelar la media y las covariables de manera directa bajo una especificación de modelos lineales generalizados. Así, si se utiliza el enlace logístico, los parámetros pueden ser interpretados en términos del ratio de odds.

Sin embargo, la distribución beta no permite una mayor flexibilidad en la especificación de la varianza y no considera eventos de colas pesadas. En este sentido, [Bayes et al. \(2012\)](#) consideraron la distribución beta rectangular, desarrollada por [Hahn \(2008\)](#) y cuya función de densidad es definida por:

$$r(y | \mu, \phi, \theta) = \theta + (1 - \theta) b(y | \mu, \phi), \quad y \in (0, 1),$$

donde $\mu \in (0, 1)$ es la media de la respuesta, $\phi > 0$ es el parámetro de dispersión, $\theta \in [0, 1]$ es el parámetro de mixtura que simplifica la distribución a una uniforme cuando $\theta = 1$ y a una beta cuando $\theta = 0$, y con media igual a $E(y) = \theta/2 + (1 - \theta)\mu$. Del mismo modo, los autores realizaron una reparametrización de la distribución considerando $\gamma = \theta/2 + (1 - \theta)\mu$ y $\alpha = \frac{\theta}{1 - (1 - \theta)|2\mu - 1|}$ con el fin de obtener una estructura de regresión para la media más adecuada, la cual será denota por $Y \sim BRR(\gamma, \phi, \alpha)$, y cuya función de densidad viene dada por:

$$r(y | \gamma, \phi, \alpha) = \alpha(1 - |2\gamma - 1|) + (1 - \alpha(1 - |2\gamma - 1|)) b\left(y \mid \frac{\gamma - 0,5\alpha(1 - |2\gamma - 1|)}{1 - \alpha(1 - |2\gamma - 1|)}, \phi\right), \tag{1.6}$$

donde $y \in (0, 1)$, $\gamma \in [0, 1]$ viene a ser la media de la variable respuesta, $\alpha \in [0, 1]$ es un parámetro de forma que está asociado con el ancho de las colas de la distribución, y $\phi > 0$ es el parámetro de dispersión. En este sentido, si variable y_i sigue esta última distribución,

el modelo de regresión beta rectangular se puede representar de la siguiente manera:

$$\begin{aligned} Y_i &\sim BRR(\gamma_i, \phi_i, \alpha), \\ g_1(\gamma_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \gamma_i \in (0, 1), \\ g_2(\phi_i) &= -\mathbf{w}_i^\top \boldsymbol{\eta} + \epsilon_i, \quad \phi_i > 0, \end{aligned} \quad (1.7)$$

donde γ_i es la media de la respuesta; $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_{k_1}]^\top$ y $\boldsymbol{\eta} = [\eta_0, \eta_1, \dots, \eta_{k_2}]^\top$ son vectores columna de parámetros; $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ik_1}]^\top$ y $\mathbf{w}_i = [1, w_{i1}, \dots, w_{ik_2}]^\top$ son vectores columna de covariables por individuo i ; g_1 y g_2 son funciones de enlaces estrictamente monótonas y diferenciables de segundo orden de $(0, 1)$ a \mathbb{R} para el primer caso y de \mathbb{R}^+ a \mathbb{R} para el segundo; y ϕ_i es el parámetro que controla la dispersión de la distribución. Así, esta estructura permite variar las cantidades de dispersión e incluir una mayor probabilidad de eventos extremos en las colas y así ser más robusta a la presencia de valores atípicos.

Otra distribución adecuada para modelar variables continuas limitadas en un intervalo $(0, 1)$ es la distribución simplex, desarrollada por [Barndorff-Nielsen y Jorgensen \(1991\)](#). Al respecto, una variable aleatoria continua Y definida en el intervalo abierto $(0, 1)$ tiene distribución simplex con media μ y parámetro de dispersión σ^2 , y que se denota por $Y \sim S^-(\mu, \sigma^2)$, si su función de densidad de probabilidad es dada por:

$$f(y | \mu, \sigma^2) = \left[2\pi\sigma^2 \{y(1-y)\}^3 \right]^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} d(y | \mu) \right\}, \quad y \in (0, 1). \quad (1.8)$$

En este sentido, si la variable y_i sigue esta última distribución, el modelo de regresión simplex se puede representar de la siguiente manera:

$$\begin{aligned} Y_i &\sim S^-(\mu_i, \sigma^2), \\ g(\mu_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad \mu_i \in (0, 1), \end{aligned} \quad (1.9)$$

donde μ_i es la media de la respuesta; $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ik}]^\top$ son vectores columna de covariables por individuo i ; $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_k]^\top$ son vectores columna de parámetros, y g es una función de enlace estrictamente monótona y diferenciable de segundo orden de $(0, 1)$ a \mathbb{R} .

Cabe señalar que, esta distribución ha sido ampliamente utilizada para modelar proporciones, por ejemplo, [Jorgensen \(1997\)](#) señala que tiene la virtud de ser un modelo de dispersión y por tanto, el análisis de desvío desarrollado para modelos lineales generalizados puede ser aplicado a modelos de regresión basados en esta distribución. Así, [Song y Tan \(2000\)](#) propusieron un modelo de regresión simplex con dispersión constante para modelar datos longitudinales bajo el enfoque de ecuaciones de estimación generalizadas, el cual luego fue ampliado por [Song et al. \(2004\)](#) al asumir que el parámetro de dispersión varía a lo largo de las observaciones. Asimismo, [Espinheira y de Oliveira Silva \(2018\)](#) propusieron el modelo de regresión simplex no lineal, que considera estructuras no lineales en los parámetros para las regresiones de la media y la dispersión. Por otro lado, [Zhang et al. \(2016\)](#) implementó el paquete `simplexreg` en R, que proporciona el ajuste del modelo de regresión simplex para

modelar proporciones.

En línea con lo anterior, [Kieschnick y McCullough \(2003\)](#) señalan que los modelos más adecuados para modelar proporciones son los modelos de regresión simplex y beta, recomendando el uso de este último. Del mismo modo, [Lopez \(2013\)](#) menciona que el modelo de regresión beta es el más apropiado para ser comparado con el modelo de regresión simplex porque permite modelar la dispersión individual de los datos. Al respecto, el autor concluyó que el modelo simplex obtiene estimaciones más cercanas de los parámetros verdaderos en relación al modelo beta. De igual forma, [Espinheira y de Oliveira Silva \(2018\)](#) compararon el ajuste de ambos modelos para dos datos reales considerando linealidad y no linealidad en los parámetros. Al respecto, los autores evidenciaron que el modelo de regresión simplex resulta ser la mejor opción, sobre todo porque la estimación por máxima verosimilitud del mismo resultó ser más robusto en comparación del modelo beta, principalmente cuando se presentan datos influyentes en los datos. Por otro lado, [Miyashiro \(2008\)](#) señala que los modelos más recomendados para el análisis de variables restringidas en el intervalo $(0, 1)$ se basan en las distribuciones beta y simplex. Al respecto, el autor propone unas medidas de diagnóstico y realiza comparaciones con datos reales para estimar los parámetros bajo el modelo de regresión simplex y beta, obteniendo estimaciones muy similares para los submodelos de la media. Finalmente, [Bandyopadhyay et al. \(2017\)](#) combinaron las distribuciones beta, beta rectangular y simplex, en una nueva clase de densidad llamada densidad de proporción general (GPD, por sus siglas en inglés), donde los tres modelos mencionados resultan ser casos particulares. Al respecto, los autores concluyeron que la densidad simplex es la más flexible y permite capturar diversas formas de densidad de datos de proporciones, incluso en situaciones donde la densidad beta puede estar lejos de la evidencia empírica (por ejemplo, cuando el parámetro de dispersión es pequeño). Por estos motivos, esta tesis considerará el modelo de regresión simplex para el análisis de proporciones y comparará sus resultados con los obtenidos bajo el modelo de regresión beta.

Otro enfoque para analizar variables que son proporciones son los modelos de regresión cuasi-paramétricos en donde solo se especifica el primer y segundo momento de la distribución condicional, pero no se especifica la distribución total. Tal como lo señala [Kieschnick y McCullough \(2003\)](#), [Cox \(1996\)](#) estudió el uso de dos funciones de enlaces (logit y log-log) bajo dos especificaciones de la función de varianza (ortogonal y canónica), para una muestra de datos limitados en un intervalo $(0, 1)$, concluyendo que la mejor combinación para la estimación de proporciones resulta ser aquel con la función de enlace logit y la función de varianza ortogonal.

Al respecto, si bien los modelos mencionados previamente son adecuados para el análisis de proporciones, existen situaciones donde se requieren medir probabilidades no nulas para los valores extremos 0 y 1, es decir, variables definidas en el intervalo cerrado $[0, 1]$. Por ejemplo, al analizar el porcentaje del uso de la línea de la tarjeta de crédito de un cliente, este puede tomar el valor de 0 cuando el cliente no ha realizado ningún consumo con la tarjeta, o el valor de 1 cuando ha consumido la línea total asignada. Otro ejemplo, que será evaluado en el presente trabajo, es analizar el grado de la dolarización de los préstamos de las empresas, que puede tomar el valor de 0 cuando solo registra deudas en moneda nacional

y de 1 cuando solo registra deudas en moneda extranjera.

Una alternativa que considera probabilidades no nulas de alcanzar los valores extremos 0 y 1 son los modelos basados en transformaciones, que buscan transformar una variable limitada cerrada a una limitada abierta, para luego aplicar cualquier modelo de regresión para intervalos abiertos, como los desarrollados anteriormente. Al respecto, [Smithson y Verkuilen \(2006\)](#) utilizaron la siguiente transformación:

$$y^* = \frac{y(N-1) + 1/2}{N}, \quad y \in [0, 1], \quad (1.10)$$

donde $y^* \in (0, 1)$ y N es el tamaño de la muestra. Sin embargo, [Galvis et al. \(2014\)](#) señalan que conviene utilizar esta alternativa solo cuando se registran pequeñas proporciones de ceros y unos en la muestra de datos, caso contrario, se incrementa la sensibilidad en la estimación de los parámetros. Asimismo, [Bayes y Valdivieso \(2016\)](#) mostraron que los estimadores de este modelo resultaron ser extremadamente sesgados e inconsistentes.

Una segunda estrategia consiste en utilizar los modelos denominados en dos etapas, que plantean ajustar un modelo de regresión binaria para estimar las probabilidades de que $Y = 0$ o $Y = 1$, para luego aplicar un modelo de regresión para intervalos abiertos para $Y \in (0, 1)$. Bajo este enfoque, se asume que los factores que influyen en la respuesta para que registre los valores extremos (0 o 1) podrían ser distintos a aquellos que determinan el valor dentro del intervalo (0, 1). Por ejemplo, [Ramalho y da Silva \(2009\)](#) propusieron el modelo cero-inflacionado para analizar los factores que determinan la elección de tener o no tener deudas por parte de una empresa, así como de estimar los niveles de deudas alcanzados para aquellos que optaron por tener deudas, para lo cual definieron la siguiente variable:

$$Y^* = \begin{cases} 0, & \text{si } Y = 0, \\ 1, & \text{si } Y \in (0, 1). \end{cases} \quad (1.11)$$

Así, en un primer paso, los autores estiman la probabilidad de que la empresa no genere deudas mediante un modelo de regresión logístico, para luego estimar los niveles de deudas alcanzados por las empresas mediante un modelo de regresión beta. Es decir, analizan la influencia de las covariables sobre el nivel de deuda independientemente de la elección previa de la empresa para generar deudas o no.

Posteriormente, [Ospina y Ferrari \(2010\)](#) ampliaron el alcance de la metodología anterior al ajustar un modelo de regresión normal multivariado para estimar las probabilidades de que $Y = 0$, $Y = 1$ y $Y \in (0, 1)$ como variable dependiente. Así, propusieron los modelos de regresión beta inflacionada en unos para intervalos (0, 1], beta inflacionada en ceros para intervalos [0, 1) y beta inflacionada en ceros y unos para intervalos [0, 1]. Esta última distribución es esencialmente una mixtura entre una distribución Bernoulli y una beta y tiene la

siguiente función de masa:

$$f_Y(y|\delta_0, \delta_1, \mu, \phi) = \begin{cases} \delta_0 & , \text{ si } y = 0, \\ (1 - \delta_0 - \delta_1) b(y|\mu, \phi) & , \text{ si } y \in (0, 1), \\ \delta_1 & , \text{ si } y = 1, \end{cases} \quad (1.12)$$

donde $\delta_0 \in (0, 1)$ denota la probabilidad que Y tome el valor 0, $\delta_1 \in (0, 1)$ la probabilidad que Y tome el valor 1, $(1 - \delta_0 - \delta_1) \in (0, 1)$ la probabilidad que Y tome valores en el intervalo $(0, 1)$, y b es la función de densidad de probabilidad de una distribución beta reparametrizada dada en (1.4).

Sin embargo, una debilidad en estos modelos es que se torna más difícil la interpretación del efecto de las covariables sobre la media de la respuesta al depender de varios parámetros. Por ejemplo, para la distribución beta inflacionada en 0 y 1, el valor esperado de Y es $E(Y) = \delta_1 + (1 - \delta_0 - \delta_1)\mu$. Ante este problema, [Bayes y Valdivieso \(2016\)](#) proponen una reparametrización alternativa para la distribución beta inflacionada que permite una interpretación más directa de la media, la cual será denotada por $Y_i \sim BIm(\alpha_0, \alpha_1, \gamma, \phi)$ y cuya función de masa de probabilidad viene dada por:

$$B_Y(y|\alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} \alpha_0(1 - \gamma) & , \text{ si } y = 0, \\ (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma) b\left(y \left| \frac{\gamma(1-\alpha_1)}{1-\alpha_0(1-\gamma)-\alpha_1\gamma}, \phi \right.\right) & , \text{ si } y \in (0, 1), \\ \alpha_1\gamma & , \text{ si } y = 1, \end{cases} \quad (1.13)$$

donde $\alpha_0 \in (0, 1)$, $\alpha_1 \in (0, 1)$, $\gamma \in (0, 1)$, $\phi > 0$, y b es la función de densidad de probabilidad de una distribución beta reparametrizada dada en (1.4). Asimismo, los autores proponen ajustar un conjunto de ecuaciones de regresión para estimar la media de la variable respuesta $\gamma = \delta_1 + (1 - \delta_0 - \delta_1)\mu$ y los parámetros $\alpha_0 = \frac{\delta_0}{1-\gamma}$ y $\alpha_1 = \frac{\delta_1}{\gamma}$, lo cual lo denominaron el modelo de regresión a la media beta inflacionada. Considerando esta metodología, si se tiene variables aleatorias limitadas Y_1, \dots, Y_n que siguen esta última distribución, la formulación del modelo de regresión a la media beta inflacionada es la siguiente:

$$\begin{aligned} Y_i &\sim BIm(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi), \\ g_1(\alpha_{0i}) &= \hat{\mathbf{x}}_i^\top \boldsymbol{\omega} + \epsilon_i, \\ g_2(\alpha_{1i}) &= \check{\mathbf{x}}_i^\top \boldsymbol{\eta} + \epsilon_i, \\ g_3(\gamma_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \end{aligned} \quad (1.14)$$

donde $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_{k_1}]^\top$, $\boldsymbol{\eta} = [\eta_0, \eta_1, \dots, \eta_{k_2}]^\top$ y $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_{k_3}]^\top$ son vectores columna de parámetros de regresión; $\hat{\mathbf{x}}_i = [1, \hat{x}_{i1}, \dots, \hat{x}_{ik_1}]^\top$, $\check{\mathbf{x}}_i = [1, \check{x}_{i1}, \dots, \check{x}_{ik_2}]^\top$ y $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ik_3}]^\top$ son vectores columna de covariables por individuo i ; g_1 , g_2 y g_3 son funciones de enlace estrictamente monótonas y diferenciables de segundo orden de $(0, 1)$ a \mathbb{R} . Bajo este enfoque se puede analizar la influencia de las covariables directamente sobre la media,

a diferencia de los modelos en dos etapas que se realizaba sobre la media condicional. En este sentido, esta metodología será utilizada en esta tesis por lo que será desarrollada en mayor detalle en los siguientes capítulos y en donde en vez de utilizar la distribución beta, se trabajará con la distribución simplex.

1.2. Objetivos

Se propone el modelo de regresión a la media simplex inflacionada desde el punto de vista de la estadística clásica, desarrollando las principales propiedades, estimando los parámetros y realizando una aplicación del modelo a un conjunto de datos reales. De manera específica:

- Revisar la literatura de modelos de regresión donde la variable dependiente se encuentra limitada en un intervalo.
- Proponer el modelo de regresión a la media simplex inflacionada y analizar sus propiedades.
- Estimar los parámetros del modelo propuesto desde la perspectiva de la estadística clásica.
- Aplicar el modelo propuesto para determinar el grado de dolarización de empresas que registran deudas en el Sistema Financiero bajo la denominación de créditos corporativos, a grandes empresas y a medianas empresas.

1.3. Organización del Trabajo

En el Capítulo 2, se presenta la distribución simplex, que permite modelar variables continuas limitadas en el intervalo abierto $(0, 1)$, y la distribución simplex inflacionada, que considera probabilidades no nulas de alcanzar los valores extremos 0 y 1. Sobre esta última, se realiza una parametrización alternativa con el fin de obtener una interpretación más directa de la media, cuya distribución resultante será utilizada en los demás capítulos. Cabe señalar que, para cada distribución, se define su función de masa de probabilidad y las principales propiedades (esperanza y varianza).

En el Capítulo 3, se propone el modelo de regresión a la media simplex inflacionada, que permite modelar variables aleatorias continuas limitadas en el intervalo cerrado $[0, 1]$ al considerar un conjunto de ecuaciones de regresión para estimar la media de la respuesta y los parámetros que miden las probabilidades de los valores extremos 0 y 1. Asimismo, se detalla el método para la estimación de los parámetros desde la perspectiva de la estadística clásica.

En el Capítulo 4, se presenta un estudio de simulación con el fin de evaluar si el método propuesto permite recuperar los parámetros del modelo de regresión a la media simplex inflacionada. Para tal fin, los criterios que se utilizan para evaluar el desempeño de los estimadores son el sesgo relativo y el error cuadrático medio, mientras que para el desempeño de los intervalos de confianza se considera el ratio de cobertura para un nivel de significación del 95%. Además, se muestra el coeficiente de determinación R^2 para analizar el ajuste de los modelos.

En el Capítulo 5, se desarrolla la aplicación del modelo propuesto para determinar el grado de dolarización de empresas que registran deudas en el Sistema Financiero bajo la denominación de créditos corporativos, a grandes empresas y a medianas empresas. Asimismo, se ajusta los datos considerando el modelo de regresión a la media beta inflacionada, con el fin de comparar el desempeño de ambos modelos, para lo cual se ha analizado el criterio de información de akaike (AIC, por sus siglas en inglés) y el condicional (AICc), el criterio de información bayesiano (BIC, por sus siglas en inglés), el error cuadrático medio (MSE, por sus siglas en inglés) y el coeficiente de determinación R^2 .

Finalmente, en el Capítulo 6 se presentan las principales conclusiones obtenidas en esta tesis así como las sugerencias para las futuras investigaciones.

En el anexo se presenta la especificación de la función score y de la matriz hessiana del modelo (Apéndice A).



Capítulo 2

Distribución Simplex Inflacionada

Este capítulo se dedica al estudio de la distribución simplex en su especificación básica, para luego analizar la distribución simplex inflacionada, que resulta de una mezcla entre una distribución simplex y una Bernoulli. Sobre esta última se considera una parametrización alternativa, cuya distribución resultante es utilizada en los siguientes capítulos. Cabe señalar que, para cada distribución, se define su función de masa de probabilidad y las principales propiedades (esperanza y varianza).

2.1. Distribución Simplex

La distribución simplex, desarrollada por [Barndorff-Nielsen y Jorgensen \(1991\)](#), ha sido ampliamente utilizada para modelar variables continuas limitadas en el intervalo $(0, 1)$. Al respecto, [Jorgensen \(1997\)](#) señala que la distribución tiene la virtud de ser un modelo de dispersión y por tanto, el análisis de desvío desarrollado para modelos lineales generalizados puede ser aplicado a modelos de regresión basados en esta distribución.

2.1.1. Función de densidad de probabilidad

Una variable aleatoria continua W definida en el intervalo abierto $(0, 1)$ tiene distribución simplex con media μ y parámetro de dispersión σ^2 , y que se denota por $W \sim S^-(\mu, \sigma^2)$, si su función de densidad de probabilidad es dada por:

$$S_W(w | \mu, \sigma^2) = \left[2\pi\sigma^2 \{w(1-w)\}^3 \right]^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} d(w | \mu) \right\}, \quad w \in (0, 1), \quad (2.1)$$

donde d es la función de desvío, definida por:

$$d(w | \mu) = \frac{(w - \mu)^2}{w(1-w)\mu^2(1-\mu)^2}.$$

Los parámetros $\mu \in (0, 1)$ y $\sigma^2 > 0$ caracterizan la función de densidad (ver Figura 2.1), pudiendo tomar la forma de una campana, U, J o L (conocida como inversa-J). Al respecto, cuando el parámetro de dispersión $\sigma^2 \rightarrow 0$ (gráficas de la fila 4), la distribución simplex es muy similar a la distribución normal debido a la teoría asintótica ([Jorgensen, 1997](#)):

$$\frac{W - \mu}{\sigma\sqrt{V(\mu)}} \xrightarrow{d} N(0, 1),$$

donde $V(\mu) = \mu^3(1 - \mu)^3$ es la función de varianza. Por el contrario, a medida que σ^2 se incrementa, existe una mayor dispersión de los datos hacia los valores extremos con relación a la media μ , generando distribuciones asimétricas de colas a la derecha, izquierda o ambas. Asimismo, la distribución es unimodal si $\sigma \leq 4/\sqrt{3}$ (gráficas de las filas 3 y 4), caso contrario, es multimodal (gráficas de las filas 1 y 2).

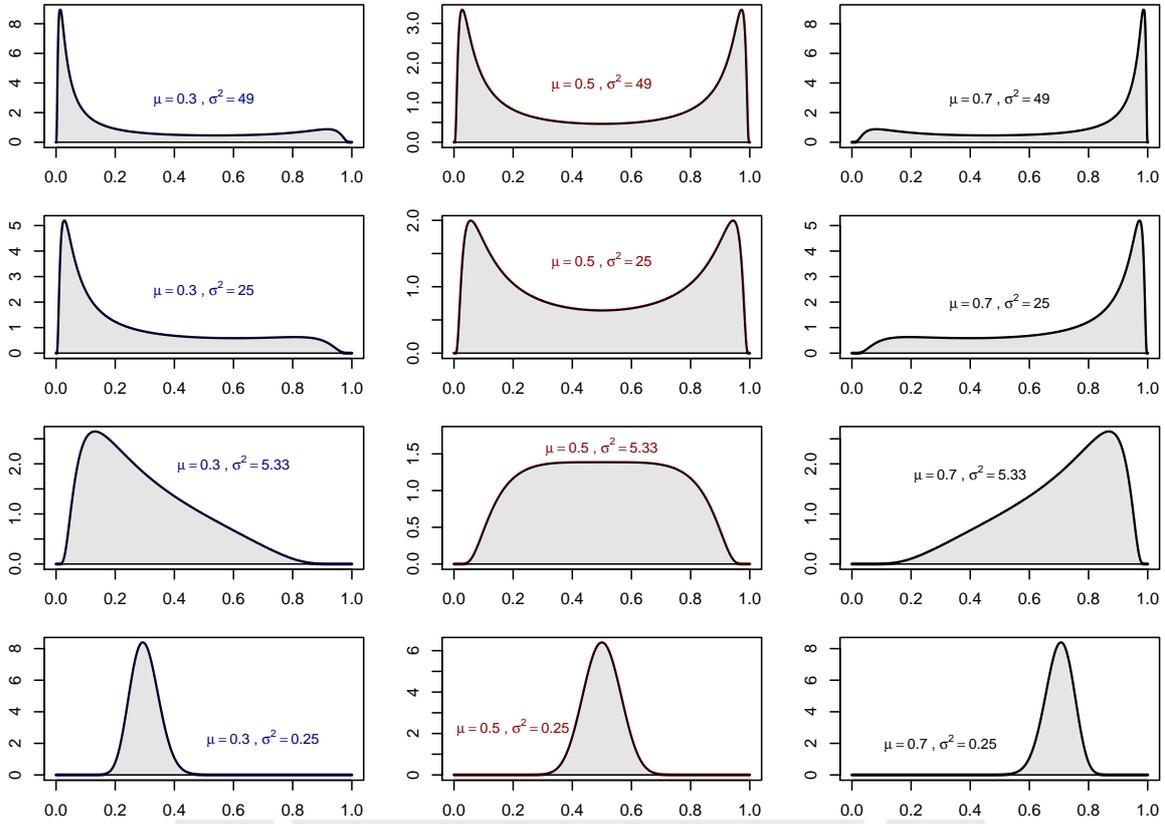


Figura 2.1: Función de densidad de probabilidad de la distribución simplex con media μ (de izquierda a derecha: 0.3,0.5,0.7) y parámetro de dispersión σ^2 (de arriba a abajo: 49,25,5.33,0.25)

2.1.2. Propiedades

La media y varianza de una variable aleatoria $W \sim S^-(\mu, \sigma^2)$ son dadas por:

$$\begin{aligned}
 E(W) &= \mu, \\
 V(W) &= \mu(1 - \mu) - \frac{1}{\sqrt{2\sigma^2}} \exp\left\{ \frac{1}{2\sigma^2\mu^2(1-\mu)^2} \right\} \Gamma\left\{ \frac{1}{2}, \frac{1}{2\sigma^2\mu^2(1-\mu)^2} \right\},
 \end{aligned} \tag{2.2}$$

donde Γ es la función gamma incompleta definida por $\Gamma(a, b) = \int_b^\infty t^{a-1} b^t dt$. La varianza de W depende de μ y del parámetro de dispersión σ^2 , tal como se observa en la Figura 2.2. Al respecto, manteniendo constante μ , la varianza se incrementa a medida que σ^2 aumenta. Por otro lado, manteniendo constante σ^2 , la varianza se incrementa hasta que $\mu = 0.5$ y para valores superiores la varianza disminuye.

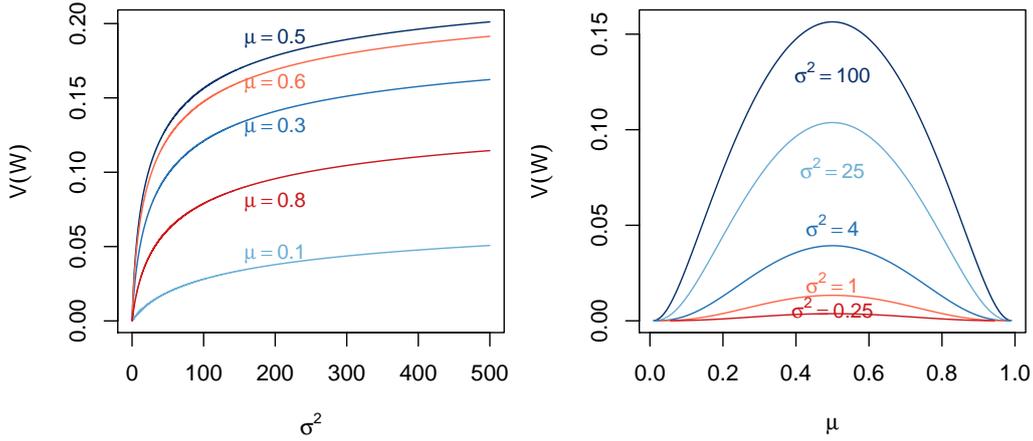


Figura 2.2: Relación de la varianza $V(W)$ contra el parámetro de dispersión σ^2 (gráfica de la izquierda, para distintos valores de $\mu = \{0.1, 0.3, 0.5, 0.6, 0.8\}$), y contra la media μ (gráfica de la derecha, para distintos valores de $\sigma^2 = \{0.25, 1, 4, 25, 100\}$), para una variable aleatoria que sigue una distribución simplex.

2.2. Distribución Simplex Inflacionada

En la Sección 2.1 se presentó la distribución simplex, que permite modelar variables limitadas en el intervalo abierto $(0, 1)$. Sin embargo, existen situaciones donde se requieren modelar probabilidades no nulas para los valores extremos 0 y 1, es decir, variables definidas en el intervalo cerrado $[0, 1]$. Por ejemplo, al analizar el porcentaje del uso de la línea de la tarjeta de crédito de un cliente, este puede tomar el valor de 0 cuando el cliente no ha realizado ningún consumo con la tarjeta, o el valor de 1 cuando ha consumido la línea total asignada. Otro ejemplo, que será evaluado en esta tesis, es analizar el grado de la dolarización de empresas que registran deudas en el Sistema Financiero, que puede registrar los valores 0 o 1 cuando el total de créditos fue otorgado en moneda nacional o en moneda extranjera, respectivamente.

Un modelo alternativo que considera probabilidades no nulas de alcanzar los valores extremos es la distribución simplex inflacionada, que resulta de una mezcla entre una distribución simplex y una distribución Bernoulli.

2.2.1. Función de masa de probabilidad

Una variable aleatoria continua Y definida en el intervalo cerrado $[0, 1]$ tiene distribución simplex inflacionada con parámetros δ_0 , δ_1 , μ y σ^2 , y que se denota por $Y \sim SI(\delta_0, \delta_1, \mu, \sigma^2)$, si su función de masa de probabilidad es:

$$SI_Y(y|\delta_0, \delta_1, \mu, \sigma^2) = \begin{cases} \delta_0 & , \text{ si } y = 0, \\ (1 - \delta_0 - \delta_1) S_W(y|\mu, \sigma^2) & , \text{ si } y \in (0, 1), \\ \delta_1 & , \text{ si } y = 1, \end{cases} \quad (2.3)$$

donde $\delta_0 \in (0, 1)$ denota la probabilidad que Y tome el valor 0, $\delta_1 \in (0, 1)$ la probabilidad

que Y tome el valor 1, $(1 - \delta_0 - \delta_1) \in (0, 1)$ la probabilidad que Y tome valores en el intervalo $(0, 1)$, y S_W es la función de densidad de probabilidad de una distribución simplex dada en (2.1).

2.2.2. Propiedades

La media y varianza de una variable aleatoria $Y \sim SI(\delta_0, \delta_1, \mu, \sigma^2)$ son dadas por:

$$\begin{aligned} E(Y) &= \delta_1 + (1 - \delta_0 - \delta_1)\mu, \\ V(Y) &= \delta_1(1 - \delta_1) + (1 - \delta_0 - \delta_1) [V(W) + (\delta_0 + \delta_1)\mu^2 - 2\mu\delta_1], \end{aligned} \quad (2.4)$$

donde $V(W)$ es la varianza de $W \sim S^-(\mu, \sigma^2)$ definida en (2.2), cuando $y \in (0, 1)$.

2.3. Distribución Simplex Inflacionada Reparametrizada

Según lo señalado por Ferrari y Cribari-Neto (2004), es más útil modelar la media de la respuesta cuando se realiza un análisis de regresión para variables limitadas en el intervalo $(0, 1)$, así como definir el modelo de tal manera que incluya un parámetro de dispersión. Sin embargo, la media de una variable aleatoria $Y \sim SI(\delta_0, \delta_1, \mu, \sigma^2)$ dada en (2.4), depende de μ , δ_0 y δ_1 , y se encuentra limitada de la siguiente manera:

$$\delta_1 < E(Y) < 1 - \delta_0.$$

Esto ocasiona que el análisis de regresión para la media sea más complejo. Al respecto, Bayes y Valdivieso (2016) propusieron una parametrización alternativa con el fin de obtener una estructura de regresión simple para $E(Y)$ y permitir que los parámetros del modelo sean estimados sin ninguna restricción. Considerando esta parametrización al modelo propuesto en (2.3), se obtiene lo siguiente:

$$\begin{aligned} \gamma &= \delta_1 + (1 - \delta_0 - \delta_1)\mu, & \gamma &\in (0, 1), \\ \alpha_0 &= \frac{\delta_0}{1 - \gamma}, & \alpha_0 &\in (0, 1), \\ \alpha_1 &= \frac{\delta_1}{\gamma}, & \alpha_1 &\in (0, 1). \end{aligned} \quad (2.5)$$

Así, la media $\gamma \in (0, 1)$ y los parámetros $\alpha_0 \in (0, 1)$ y $\alpha_1 \in (0, 1)$ no se encuentran relacionados en su espacio paramétrico y pueden ser modelados independientemente.

2.3.1. Función de masa de probabilidad

Una variable aleatoria continua Y definida en el intervalo cerrado $[0, 1]$, y que considera la parametrización dada en (2.5), tiene distribución simplex inflacionada con parámetros α_0 , α_1 , γ y σ^2 , y que se denota por $Y \sim SI_m(\alpha_0, \alpha_1, \gamma, \sigma^2)$, si su función de masa de probabilidad

es dada por:

$$f_Y(y|\alpha_0, \alpha_1, \gamma, \sigma^2) = \begin{cases} \alpha_0(1 - \gamma) & , \text{ si } y = 0, \\ (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)S_W\left(y \mid \frac{\gamma(1-\alpha_1)}{1-\alpha_0(1-\gamma)-\alpha_1\gamma}, \sigma^2\right) & , \text{ si } y \in (0, 1), \\ \alpha_1\gamma & , \text{ si } y = 1, \end{cases} \quad (2.6)$$

donde S_W es la función de densidad de probabilidad de una distribución simplex dada en (2.1), y los parámetros $\alpha_0 \in (0, 1)$, $\alpha_1 \in (0, 1)$, $\gamma \in (0, 1)$ y $\sigma^2 > 0$ caracterizan la función (ver Figura 2.3).

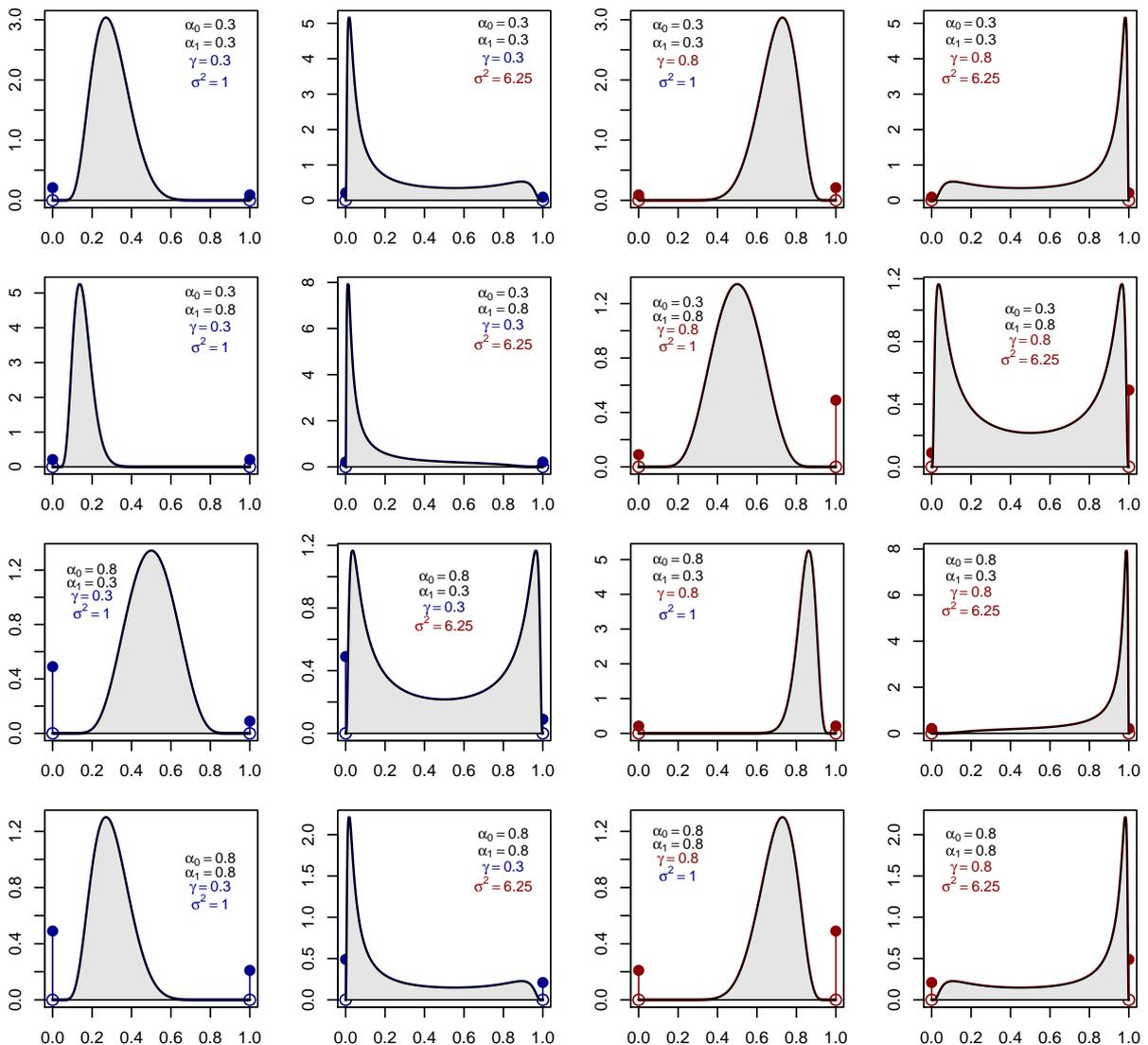


Figura 2.3: Función de densidad de probabilidad de la distribución simplex inflacionada reparametrizada con media γ (de izquierda a derecha: 0.3,0.3,0.8,0.8), parámetro de dispersión σ^2 (de izquierda a derecha: 1,6.25,1,6.25), y parámetros α_0 (de arriba a abajo: 0.3,0.3,0.8,0.8) y α_1 (de arriba a abajo: 0.3,0.8,0.3,0.8).

Al respecto, γ determina la posición de la distribución a lo largo de los valores de la misma, y σ^2 determina el grado de dispersión de los datos: cuando σ^2 se incrementa, existe una mayor dispersión de los datos hacia los valores extremos con relación a la media γ , generando distribuciones asimétricas de colas a la derecha, izquierda o ambas. Asimismo, α_0 y α_1 determinan la probabilidad que Y tome el valor de 0 y 1, respectivamente. Sobre el particular, se presentan tres distribuciones particulares: a) distribución inflacionada en unos, si $\alpha_0 = 0$ y $\alpha_1 > 0$, b) distribución inflacionada en ceros, si $\alpha_1 = 0$ y $\alpha_0 > 0$, y c) distribución simplex, si $\alpha_0 = 0$ y $\alpha_1 = 0$.

2.3.2. Propiedades

La media y varianza de una variable aleatoria $Y \sim SIM(\alpha_0, \alpha_1, \gamma, \sigma^2)$ son dadas por:

$$E(Y) = \gamma,$$

$$V(Y) = \gamma(1 - \gamma) - (1 - \alpha_0(1 - \gamma) - \alpha_1\gamma) \frac{1}{\sqrt{2\sigma^2}} \exp\{\Lambda_{(\alpha_0, \alpha_1, \gamma, \sigma^2)}\} \Gamma\left\{\frac{1}{2}, \Lambda_{(\alpha_0, \alpha_1, \gamma, \sigma^2)}\right\}, \quad (2.7)$$

donde $\Lambda_{(\alpha_0, \alpha_1, \gamma, \sigma^2)} = \frac{(1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)^4}{2\sigma^2(\gamma(1 - \gamma)(1 - \alpha_0)(1 - \alpha_1))^2}$.

La varianza de Y depende de γ y del parámetro de dispersión σ^2 , tal como se observa en la Figura 2.4. Al respecto, manteniendo constante γ , la varianza se incrementa a medida que σ^2 aumenta. Por otro lado, manteniendo constante σ^2 , la varianza se incrementa hasta que $\gamma = 0.5$ y para valores superiores la varianza disminuye.

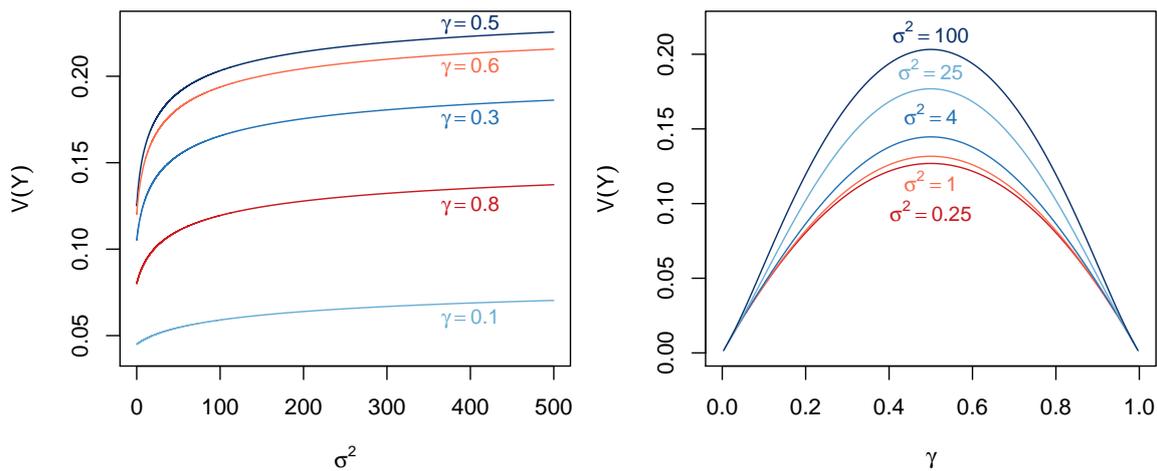


Figura 2.4: Relación de la varianza $V(Y)$ contra el parámetro de dispersión σ^2 (gráfica de la izquierda, para distintos valores de $\gamma = \{0.1, 0.3, 0.5, 0.6, 0.8\}$), y contra la media γ (gráfica de la derecha, para distintos valores de $\sigma^2 = \{0.25, 1, 4, 25, 100\}$), para una variable aleatoria que sigue una distribución simplex inflacionada reparametrizada.

Capítulo 3

Modelo de Regresión a la Media Simplex Inflacionada

Este capítulo propone el modelo de regresión a la media simplex inflacionada, que permite modelar variables aleatorias continuas limitadas en el intervalo cerrado $[0, 1]$ al considerar un conjunto de ecuaciones de regresión para estimar la media de la respuesta y los parámetros que modelan las probabilidades de los valores extremos 0 y 1. Asimismo, se detalla el método para la estimación de los parámetros desde la perspectiva de la estadística clásica.

3.1. Especificación del modelo de regresión a la media simplex inflacionada

El modelo probabilístico de la distribución simplex inflacionada dado en (2.6) permite una estructura simple para un análisis de regresión de la media γ , la cual no se encuentra relacionada con los parámetros α_0 y α_1 en su espacio paramétrico. En este sentido, considerando la metodología desarrollada por [Bayes y Valdivieso \(2016\)](#), se establecen un conjunto de ecuaciones de regresión para estimar la media de la respuesta γ y los parámetros α_0 y α_1 que modelan las probabilidades de los valores extremos 0 y 1, respectivamente, obteniendo así el modelo de regresión a la media simplex inflacionada. Bajo este enfoque, se puede analizar la influencia de las covariables directamente sobre la media, a diferencia de los modelos de regresión en dos etapas que se realizaba sobre la media condicional. Así, si se tiene variables aleatorias limitadas Y_1, \dots, Y_n que siguen una distribución simplex inflacionada, la especificación del modelo es la siguiente:

$$\begin{aligned} Y_i &\sim SIM(\alpha_{0i}, \alpha_{1i}, \gamma_i, \sigma^2), \\ g_1(\alpha_{0i}) &= \hat{\mathbf{x}}_i^\top \boldsymbol{\omega} + \epsilon_i, \\ g_2(\alpha_{1i}) &= \check{\mathbf{x}}_i^\top \boldsymbol{\eta} + \epsilon_i, \\ g_3(\gamma_i) &= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \end{aligned} \tag{3.1}$$

donde $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_{k_1}]^\top$, $\boldsymbol{\eta} = [\eta_0, \eta_1, \dots, \eta_{k_2}]^\top$ y $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_{k_3}]^\top$ son vectores columna de parámetros; $\hat{\mathbf{x}}_i = [1, \hat{x}_{i1}, \dots, \hat{x}_{ik_1}]^\top$, $\check{\mathbf{x}}_i = [1, \check{x}_{i1}, \dots, \check{x}_{ik_2}]^\top$ y $\mathbf{x}_i = [1, x_{i1}, \dots, x_{ik_3}]^\top$ son vectores columna de covariables por individuo i ; g_1 , g_2 y g_3 son funciones de enlaces estrictamente monótonas y diferenciables de segundo orden de $(0, 1)$ a \mathbb{R} ; σ^2 es el parámetro de dispersión y los parámetros α_{0i} , α_{1i} y γ_i fueron definidos en (2.5).

La estimación de los parámetros $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma^2]^\top$ se realiza desde la perspectiva de la estadística clásica, mediante el método de máxima verosimilitud.

3.1.1. Función de verosimilitud

Sean Y_1, Y_2, \dots, Y_n variables aleatorias limitadas en el intervalo $[0, 1]$ que siguen una distribución simplex inflacionada $Y_i \sim SIM(\alpha_{0i}, \alpha_{1i}, \gamma_i, \sigma^2)$ y con valores observados y_1, y_2, \dots, y_n , entonces la función de masa de probabilidad de Y_i por individuo i dada en (2.6) puede ser reescrita de la siguiente manera:

$$f_{Y_i}(y_i | \alpha_{0i}, \alpha_{1i}, \gamma_i, \sigma^2) = \begin{cases} \alpha_{0i}(1 - \gamma_i) & , \text{ si } y_i = 0, \\ c_i S_W(y_i | \mu_i, \sigma^2) & , \text{ si } y_i \in (0, 1), \\ \alpha_{1i}\gamma_i & , \text{ si } y_i = 1, \end{cases} \quad (3.2)$$

donde

$$\begin{aligned} \alpha_{0i} &= g_1^{-1}(\hat{\boldsymbol{x}}_i^\top \boldsymbol{\omega}), \\ \alpha_{1i} &= g_2^{-1}(\hat{\boldsymbol{x}}_i^\top \boldsymbol{\eta}), \\ \gamma_i &= g_3^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta}), \\ c_i &= 1 - \alpha_{0i}(1 - \gamma_i) - \alpha_{1i}\gamma_i, \\ \mu_i &= \gamma_i(1 - \alpha_{1i})/c_i. \end{aligned}$$

Por tanto, la función de verosimilitud, que depende del vector de parámetros desconocidos $\boldsymbol{\theta}$, viene dada por:

$$L(\boldsymbol{\theta}) = f_Y(Y_1 = y_1, \dots, Y_n = y_n | \boldsymbol{\theta}) = \prod_{i=1}^n f_{Y_i}(y_i | \boldsymbol{\theta}).$$

Asumiendo, sin pérdida de generalidad, que los valores observados de la variable dependiente se encuentran ordenados según los valores 0, 1 y (0, 1)-valor, con n_0, n_1 y $n - m$ términos ($m = n_0 + n_1$), respectivamente, se puede expresar la función de verosimilitud de la siguiente manera:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_0} \alpha_{0i}(1 - \gamma_i) \prod_{i=n_0+1}^m \alpha_{1i}\gamma_i \prod_{i=m+1}^n c_i S_W(y_i | \mu_i, \sigma^2).$$

3.1.2. Estimador de máxima verosimilitud

El estimador de máxima verosimilitud (EMV) para $\boldsymbol{\theta}$ es el valor de $\hat{\boldsymbol{\theta}}$ que maximiza la verosimilitud $L(\boldsymbol{\theta})$. Sin embargo, dado que el logaritmo es una función estrictamente positiva,

la maximización se realiza sobre la función de log-verosimilitud definida por:

$$K(\boldsymbol{\theta}) = \sum_{i=1}^{n_0} \log(\alpha_{0i}(1-\gamma_i)) + \sum_{i=n_0+1}^m \log(\alpha_{1i}\gamma_i) + \sum_{i=m+1}^n \log(c_i) + \sum_{i=m+1}^n \log(S_W(y_i | \mu_i, \sigma^2)), \quad (3.3)$$

donde

$$\log(S_W(y_i | \mu_i, \sigma^2)) = -\frac{1}{2} \log \left[2\pi\sigma^2 \{y_i(1-y_i)\}^3 \right] - \frac{1}{2\sigma^2} \frac{(y_i - \mu_i)^2}{y_i(1-y_i)\mu_i^2(1-\mu_i)^2}.$$

La maximización de la log-verosimilitud se realiza mediante métodos de optimización numérica utilizando el programa R; sin embargo, la especificación de la función score y de la matriz hessiana se muestra desarrollada en el apéndice A. Cabe señalar que, de acuerdo con Ferrari y Cribari-Neto (2004), se utiliza como valor inicial el estimador $\tilde{\boldsymbol{\beta}}$ obtenido del análisis de regresión lineal de las variables respuestas transformadas $g_3(Y_1), g_3(Y_2), \dots, g_3(Y_n)$ con las covariables \mathbf{x}_i^\top , considerando solo los valores continuos de Y en el intervalo abierto $(0, 1)$. Con ello, se calcula $\tilde{\gamma}_i = g_3^{-1}(\mathbf{x}_i^\top \tilde{\boldsymbol{\beta}})$. Asimismo, para el parámetro de dispersión se considera el siguiente estimador inicial:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n d(y_i | \mu_i).$$

Por otro lado, considerando la propuesta de Bayes y Valdivieso (2016), los estimadores iniciales $\tilde{\boldsymbol{\omega}}$ y $\tilde{\boldsymbol{\eta}}$ se obtienen de la siguiente manera: 1) se estiman los parámetros $\tilde{\delta}_{0i}$ y $\tilde{\delta}_{1i}$ mediante modelos de regresión normal multivariada con enlace logístico, 2) considerando $\tilde{\gamma}_i$ y los resultados en 1), se calculan $g_1(\tilde{\delta}_{0i}/(1-\tilde{\gamma}_i))$ y $g_2(\tilde{\delta}_{1i}/\tilde{\gamma}_i)$, para luego estimar los modelos de regresión con las covariables $\hat{\mathbf{x}}_i^\top$ para el primer caso y con $\tilde{\mathbf{x}}_i^\top$ para el segundo.

Así, una vez obtenido el EMV de $\hat{\boldsymbol{\theta}}$, fácilmente se puede obtener, para cada individuo i , el EMV de las probabilidades 0, 1, y la media condicional de acuerdo a lo siguiente:

$$\begin{aligned} \hat{\delta}_{0i} &= \hat{\alpha}_{0i}(1-\hat{\gamma}_i), \\ \hat{\delta}_{1i} &= \hat{\alpha}_{1i}\hat{\gamma}_i, \\ \hat{\mu}_i &= \frac{\hat{\gamma}_i - \hat{\delta}_{1i}}{1 - \hat{\delta}_{0i} - \hat{\delta}_{1i}}, \end{aligned}$$

donde $\hat{\alpha}_{0i} = g_1^{-1}(\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\omega}})$, $\hat{\alpha}_{1i} = g_2^{-1}(\tilde{\mathbf{x}}_i^\top \hat{\boldsymbol{\eta}})$ y $\hat{\gamma}_i = g_3^{-1}(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$.

Capítulo 4

Estudio de Simulación

Este capítulo presenta un estudio de simulación con el fin de evaluar si el método descrito en el Capítulo 3 permite recuperar los parámetros del modelo de regresión a la media simplex inflacionada. Para tal fin, los criterios que se utilizan para evaluar el desempeño de los estimadores son el sesgo relativo y el error cuadrático medio, mientras que para el desempeño de los intervalos de confianza se considera el ratio de cobertura para un nivel de significancia del 95 %. Además, se muestra el coeficiente de determinación R^2 para analizar el ajuste de los modelos.

4.1. Consideraciones para la simulación

Se realizó 1000 réplicas para $n = 500, 1000$ y 2000 valores simulados de la variable dependiente que sigue una distribución simplex inflacionada reparametrizada $Y \sim SIm(\alpha_0, \alpha_1, \gamma, \sigma^2)$, cuya función de masa de probabilidad fue definida en (2.6).

En primer lugar, se simula los valores de Y . Para tal fin, se asume tres covariables para cada ecuación de regresión que modelan la media de la respuesta γ y los parámetros α_0 y α_1 . En este sentido, las covariables $\hat{\mathbf{x}}_i^\top$, $\check{\mathbf{x}}_i^\top$ y \mathbf{x}_i^\top son generadas a partir de una distribución normal multivariada con vector de medias ceros y la siguiente matrix de varianza-covarianza:

$$\Sigma = \begin{bmatrix} 1 & -0.6 & 0.3 \\ -0.6 & 1 & 0.2 \\ 0.3 & 0.2 & 1 \end{bmatrix}.$$

Luego, se establecen los parámetros iniciales $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma^2]^\top$. Al respecto, dado que cada ecuación presenta tres covariables, se tendría un total de 13 parámetros:

$$\boldsymbol{\theta} = [\omega_0, \omega_1, \omega_2, \omega_3, \eta_0, \eta_1, \eta_2, \eta_3, \beta_0, \beta_1, \beta_2, \beta_3, \sigma^2]^\top.$$

Cabe señalar que, con el fin de obtener muestras diferenciadas en relación a la cantidad de valores 0 y 1 simulados, se consideró tres escenarios, de acuerdo a lo siguiente:

(a) Escenario 1: Mayor proporción de valores 1

$$\boldsymbol{\theta}_1 = [-0.9, -1.1, +0.4, -1.6, +1.4, -0.6, -0.2, +1.8, +1.4, -0.2, -0.4, +0.3, 4]^\top.$$

Del total de valores simulados, se tiene que, en promedio, el 7.7% y 58.2% de las observaciones registran los valores extremos 0 y 1, respectivamente, mientras que el 34.1%

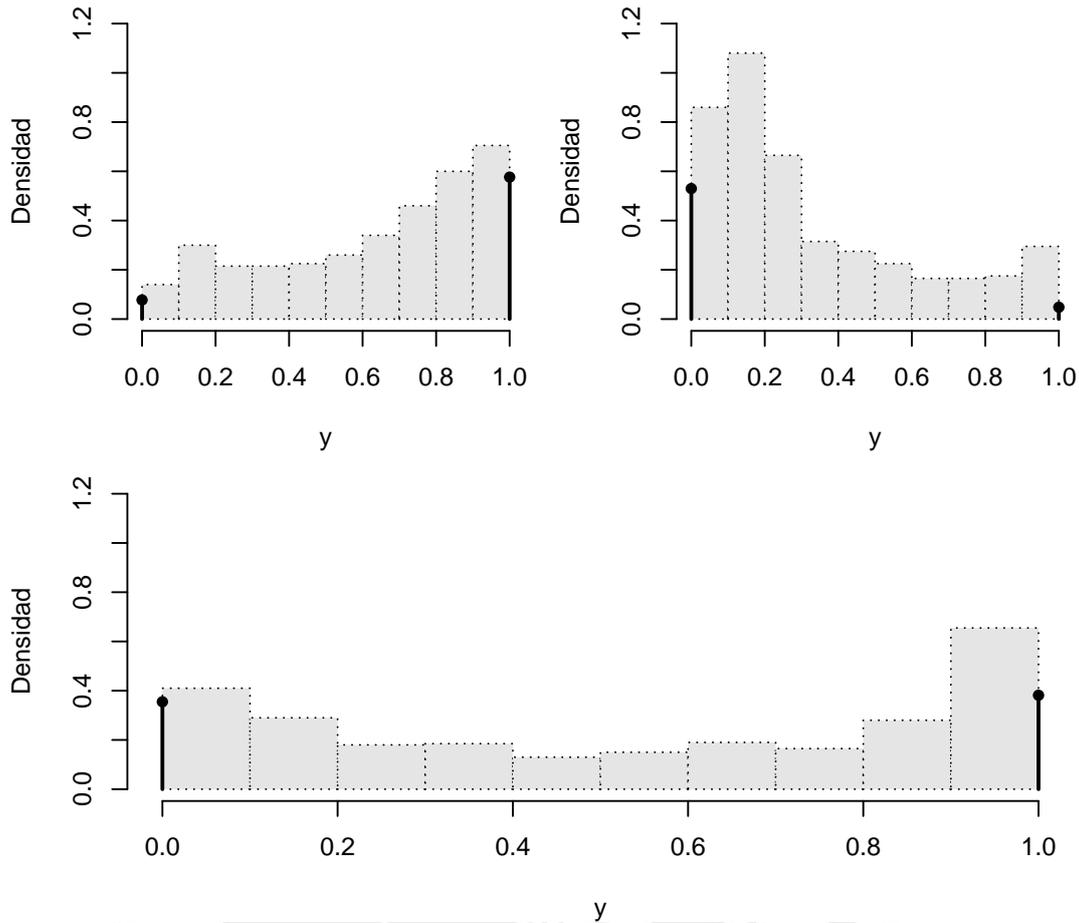


Figura 4.1: Histograma de los datos, para $n = 2000$ valores simulados de y , bajo tres escenarios distintos: a) mayor proporción de valores 1 (gráfico superior izquierdo), b) mayor proporción de valores 0 (gráfico superior derecho) y c) similar proporción de valores 0 y 1 (gráfico inferior).

registran valores en el intervalo $(0, 1)$. A modo de ejemplo, el gráfico superior izquierdo de la Figura 4.1 muestra el histograma de los datos para $n = 2000$ valores simulados de y .

- (b) Escenario 2: Mayor proporción de valores 0

$$\boldsymbol{\theta}_2 = [+1.0, +1.0, -0.5, +1.5, -1.5, +0.5, +0.3, -1.9, -1.5, +0.1, +0.5, -0.2, 4]^\top.$$

Del total de valores simulados, se tiene que en promedio el 51.7% y 5.0% de las observaciones registran los valores extremos 0 y 1, respectivamente, mientras que el 43.3% registran valores en el intervalo $(0, 1)$. A modo de ejemplo, el gráfico superior derecho de la Figura 4.1 muestra el histograma de los datos para $n = 2000$ valores simulados de y .

- (c) Escenario 3: Similar proporción de valores 0 y 1

$$\boldsymbol{\theta}_3 = [+1.5, +0.5, +0.3, -1.9, +1.5, -0.5, -0.3, +1.9, +0.1, +0.5, -0.5, -0.3, 4]^\top.$$

Del total de valores simulados, se tiene que en promedio el 35.2% y 38.4% de las observaciones registran los valores extremos 0 y 1, respectivamente, mientras que el 26.4%

registran valores en el intervalo $(0, 1)$. A modo de ejemplo, el gráfico inferior de la Figura 4.1 muestra el histograma de los datos para $n = 2000$ valores simulados de y .

Así, siendo conocidos los parámetros y covariables, para cada escenario, se calculan los parámetros $\alpha_{0i} = g_1^{-1}(\hat{\mathbf{x}}_i^\top \boldsymbol{\omega})$, $\alpha_{1i} = g_2^{-1}(\check{\mathbf{x}}_i^\top \boldsymbol{\eta})$ y $\gamma_i = g_3^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$, que en conjunto permiten simular la variable respuesta Y .

En segundo lugar, se evalúa si el método descrito en (3.1) permite recuperar los parámetros del modelo de regresión a la media simplex inflacionada. Para tal fin, considerando la función de enlace logística $g(x) = \log\left(\frac{x}{1-x}\right)$ que va de $(0, 1)$ a \mathbb{R} , se ajusta un conjunto de regresiones con las covariables $\hat{\mathbf{x}}_i^\top$, $\check{\mathbf{x}}_i^\top$ y \mathbf{x}_i^\top para estimar los parámetros $\hat{\alpha}_{0i}$ y $\hat{\alpha}_{1i}$ y la media de la respuesta $\hat{\gamma}_i$, de acuerdo a lo siguiente:

$$\begin{aligned} g_1(\alpha_{0i}) &= \omega_0 + \omega_1 \hat{x}_{1i} + \omega_2 \hat{x}_{2i} + \omega_3 \hat{x}_{3i}, \\ g_2(\alpha_{1i}) &= \eta_0 + \eta_1 \check{x}_{1i} + \eta_3 \check{x}_{2i} + \eta_3 \check{x}_{3i}, \\ g_3(\gamma_i) &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}. \end{aligned}$$

4.2. Resultados

En la Tabla 4.1 se muestra para cada n y parámetro $\boldsymbol{\theta} = [\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma^2]^\top$, los resultados de los tres criterios utilizados para evaluar el desempeño de los estimadores bajo el modelo de regresión a la media simplex inflacionada. Al respecto, se consideró las 1000 réplicas y la maximización de la función de log-verosimilitud se realizó con los estimadores iniciales propuestos en la Sección 3.

En relación al sesgo relativo, en la mayoría de los casos se observa que, a medida que aumenta el tamaño de la muestra, el sesgo disminuye para los tres escenarios. Sin embargo, se presenta un ligero incremento en el sesgo relativo cuando se incrementa la muestra a $n = 2000$ para los parámetros $\boldsymbol{\omega}$ en el escenario 1 y $\boldsymbol{\eta}$ en el escenario 2, es decir, los parámetros asociados para estimar las probabilidades de los valores extremos, y en donde los escenarios mostraban una mayor proporción de dichos valores.

Con respecto al error cuadrático medio (en adelante, ECM), para todos los casos se verifica que, a medida que aumenta el tamaño de la muestra, el ECM disminuye para los tres escenarios.

Finalmente, con relación al nivel de cobertura, se evidencia que, a medida que aumenta el número de observaciones, la cobertura se incrementa para los tres escenarios.

En este sentido, los resultados muestran que el algoritmo propuesto permite recuperar los parámetros del modelo de regresión a la media simplex inflacionada descrito en el Capítulo 3.

Parámetro	n	Escenario 1			Escenario 2			Escenario 3		
		Sesgo relativo	ECM	Cobertura	Sesgo relativo	ECM	Cobertura	Sesgo relativo	ECM	Cobertura
ω_0	500	-0.0038	0.0375	0.9491	0.0050	0.0102	0.9495	0.0017	0.0117	0.9565
	1000	-0.0010	0.0158	0.9630	0.0030	0.0052	0.9476	0.0027	0.0059	0.9531
	2000	-0.0003	0.0079	0.9377	0.0033	0.0026	0.9381	0.0005	0.0029	0.9504
ω_1	500	-0.0031	0.2018	0.9450	0.0139	0.1606	0.9424	-0.0122	0.1073	0.9393
	1000	0.0031	0.0790	0.9440	0.0006	0.0694	0.9486	-0.0336	0.0484	0.9388
	2000	-0.0012	0.0345	0.9410	0.0064	0.0305	0.9533	0.0064	0.0196	0.9483
ω_2	500	0.0111	0.1747	0.9460	-0.0195	0.1431	0.9374	-0.0334	0.0954	0.9372
	1000	-0.0107	0.0691	0.9388	0.0009	0.0593	0.9537	-0.0517	0.0442	0.9409
	2000	0.0031	0.0302	0.9344	-0.0133	0.0278	0.9543	0.0131	0.0176	0.9463
ω_3	500	0.0022	0.0776	0.9460	-0.0038	0.0616	0.9364	-0.0021	0.0407	0.9443
	1000	-0.0005	0.0307	0.9409	0.0023	0.0269	0.9507	-0.0052	0.0176	0.9419
	2000	0.0005	0.0127	0.9422	-0.0034	0.0114	0.9503	0.0012	0.0075	0.9463
η_0	500	0.0053	0.0093	0.9491	0.0163	0.0652	0.9626	0.0059	0.0117	0.9484
	1000	0.0015	0.0052	0.9345	0.0127	0.0322	0.9406	0.0018	0.0061	0.9429
	2000	0.0024	0.0023	0.9433	0.0041	0.0133	0.9563	0.0030	0.0028	0.9514
η_1	500	-0.0309	0.1545	0.9460	-0.0278	0.4207	0.9414	-0.0183	0.1197	0.9281
	1000	-0.0099	0.0687	0.9483	0.0099	0.1662	0.9466	-0.0102	0.0467	0.9582
	2000	0.0188	0.0315	0.9533	-0.0207	0.0786	0.9503	-0.0038	0.0227	0.9380
η_2	500	-0.0981	0.1446	0.9420	-0.0417	0.3707	0.9404	-0.0288	0.1039	0.9322
	1000	-0.0412	0.0624	0.9504	0.0286	0.1435	0.9496	-0.0133	0.0419	0.9521
	2000	0.0541	0.0307	0.9511	-0.0283	0.0693	0.9523	-0.0029	0.0200	0.9360
η_3	500	-0.0028	0.0581	0.9521	0.0064	0.1640	0.9333	-0.0014	0.0441	0.9291
	1000	-0.0014	0.0254	0.9483	0.0089	0.0656	0.9496	-0.0004	0.0173	0.9592
	2000	0.0043	0.0120	0.9488	-0.0002	0.0309	0.9523	0.0000	0.0083	0.9442
β_0	500	0.0033	0.0057	0.9460	0.0036	0.0044	0.9485	0.0347	0.0072	0.9555
	1000	0.0010	0.0032	0.9324	0.0022	0.0024	0.9406	-0.0224	0.0036	0.9429
	2000	0.0012	0.0013	0.9666	0.0010	0.0011	0.9462	0.0142	0.0018	0.9411
β_1	500	0.0660	0.0643	0.9440	0.0584	0.0637	0.9333	-0.0093	0.0568	0.9302
	1000	-0.0319	0.0290	0.9556	0.0213	0.0253	0.9496	-0.0022	0.0232	0.9337
	2000	0.0141	0.0130	0.9488	0.0141	0.0127	0.9523	0.0017	0.0098	0.9525
β_2	500	0.0349	0.0536	0.9430	0.0107	0.0520	0.9283	0.0122	0.0506	0.9332
	1000	-0.0123	0.0241	0.9546	-0.0006	0.0217	0.9587	0.0029	0.0199	0.9368
	2000	0.0090	0.0112	0.9533	0.0045	0.0110	0.9472	-0.0026	0.0088	0.9494
β_3	500	0.0351	0.0237	0.9460	0.0200	0.0225	0.9333	-0.0072	0.0220	0.9241
	1000	-0.0067	0.0108	0.9472	0.0100	0.0094	0.9567	-0.0017	0.0089	0.9409
	2000	0.0077	0.0048	0.9499	0.0031	0.0048	0.9442	0.0040	0.0037	0.9483
σ^2	500	-0.0283	0.0129	0.8870	-0.0236	0.0108	0.8727	-0.0389	0.0167	0.8644
	1000	-0.0159	0.0059	0.9197	-0.0143	0.0048	0.9204	-0.0200	0.0081	0.9011
	2000	-0.0072	0.0029	0.9433	-0.0045	0.0025	0.9442	-0.0126	0.0040	0.9225

Cuadro 4.1: Desempeño de los estimadores sesgo relativo, error cuadrático medio (ECM) y cobertura, bajo tres escenarios: 1) mayor proporción de valores unos, 2) mayor proporción de valores ceros y 3) similar proporción de valores ceros y unos, para $n = 500, 1000$ y 2000 valores simulados de y que siguen una distribución simplex inflacionada reparametrizada.

Capítulo 5

Aplicación

5.1. El riesgo cambiario crediticio

Las empresas que operan en el Sistema Financiero (en adelante, empresas financieras¹) se encuentran expuestas a una diversidad de riesgos tales como el riesgo de crédito, de liquidez, de mercado, operacional, entre otros. Al respecto, el riesgo de crédito es uno de los principales riesgos que enfrentan y que, según lo establecido en el Reglamento de Gestión de Riesgo de Crédito (aprobado por Resolución SBS N° 3780-2011), se define como la posibilidad de pérdidas por la incapacidad o falta de voluntad de los deudores, contrapartes, o terceros obligados, para cumplir sus obligaciones contractuales registradas dentro o fuera del balance. Este riesgo puede surgir por diversas fuentes (internas o externas), siendo una de ellas las pérdidas crediticias asociadas a variaciones en el tipo de cambio, lo cual se conoce como el riesgo cambiario crediticio (en adelante, RCC).

Sobre el particular, el Reglamento para la Administración del Riesgo Cambiario Crediticio (aprobado por Resolución SBS N° 41-2005) define el RCC como la posibilidad de afrontar pérdidas derivadas de incumplimientos de los deudores en el pago de sus obligaciones crediticias producto de descalces entre sus exposiciones netas en moneda extranjera. Asimismo, el artículo 3° del citado reglamento establece que las empresas financieras deberán adoptar un sistema de control de RCC para las colocaciones en moneda extranjera que identifique, mida, controle y reporte adecuadamente sus niveles de exposición, y que incluya la evaluación permanente de los mecanismos de control, así como de las acciones correctivas o mejoras requeridas, según sea el caso. Para tal fin, las empresas financieras realizan un proceso de evaluación de los deudores mediante una metodología interna a fin de establecer si se encuentran expuestos o no a dicho riesgo según los estándares propios de la empresa. Algunas de las metodologías consisten en establecer umbrales para el porcentaje de deuda en moneda extranjera respecto de la deuda total (en adelante, dolarización), la relación que debe existir entre el porcentaje de ventas en moneda extranjera y el porcentaje de deuda en moneda extranjera, así como márgenes de ventas diferenciados por sector económico al que pertenece el deudor, entre otros enfoques.

Dado que la dolarización permite medir de manera simple el grado de exposición al riesgo cambiario crediticio de los deudores, y considerando que es una variable continua limitada

¹ Las empresas financieras comprenden a las empresas bancarias, financieras, cajas municipales de ahorro y crédito (CMAC), cajas rurales de ahorro y crédito (CRAC), entidades de desarrollo de la pequeña y microempresa (Edpymes) y empresas de arrendamiento financiero.

en el intervalo $[0, 1]$ que puede tomar el valor de 0 cuando el deudor solo registra deudas en moneda nacional y de 1 cuando solo registra deudas en moneda extranjera, se modelará dicha variable bajo el modelo de regresión a la media simplex inflacionada.

5.2. Dolarización por cartera de créditos

A mayo de 2018, la cartera de créditos en moneda extranjera del Sistema Financiero alcanzó el 29.6 % del total de colocaciones (S/ 85,932 MM). El nivel de 29.6 % se explica, por un lado, por el alto grado de dolarización de la cartera de créditos no minoristas² (48.8 %), donde los créditos a corporativos y a grandes (37.2 % de participación y 52.1 % de dolarización) y los créditos a medianas empresas (14.9 % de participación y 40.6 % de dolarización), aportaron 19.4 y 6.0 puntos porcentuales a la dolarización del total de colocaciones, respectivamente. De otro lado, la cartera de créditos minoristas³ presenta un bajo grado de dolarización (8.6 %), donde los créditos de consumo (19.8 % de participación y 5.0 % de dolarización), los créditos hipotecarios para vivienda (15.3 % de participación y 17.9 % de dolarización) y los créditos a pequeñas y microempresas (12.8 % de participación y 3.0 % de dolarización) aportaron 1.0, 2.7 y 0.4 puntos porcentuales a la dolarización del total de colocaciones, respectivamente.

Los resultados muestran que la cartera de créditos de deudores corporativos, grandes empresas y medianas empresas presentan un alto grado de dolarización. En tal sentido, esta tesis se enfocará en analizar el grado de dolarización para dicho segmento de deudores.

5.3. Base de datos

Se ha considerado la información de 675 empresas que registran deudas en el Sistema Financiero y que presentan las siguientes características:

1. El 71.1 % de empresas corresponden a deudores de créditos corporativos y a grandes empresas, mientras que el 28.9 % a deudores de créditos a medianas empresas, conforme a los criterios establecidos en el Reglamento para la Evaluación y Clasificación del Deudor y la Exigencia de Provisiones, aprobado por Resolución SBS N° 11356-2008.
2. El 11.4 % de empresas registran información de estados financieros al cierre de diciembre de 2014, el 4.1 % y 14.4 % al cierre de junio y diciembre de 2015, respectivamente, el 7.9 % y 29.0 % al cierre de junio y diciembre de 2016, respectivamente, y el 33.2 % al cierre de junio de 2017.
3. El 18.2 % de empresas desarrollan el comercio como principal actividad económica, mientras que el 12.0 % y 11.4 % corresponden a empresas de los sectores servicios y construcción, respectivamente. El 58.4 % restante corresponde a empresas de diversos sectores tales como metalmecánica y maquinarias, agricultura y ganadería, textil y cuero, entre otros.

² Se entiende por deudor no minorista a aquella persona natural o jurídica que cuenta con créditos directos o indirectos clasificados como corporativos, a grandes empresas o a medianas empresas.

³ Se entiende por deudor minorista a aquella persona natural o jurídica que cuenta con créditos directos o indirectos clasificados como consumo (revolventes y no revolventes), a microempresas, a pequeñas empresas o hipotecarios para vivienda.

4. El 6.1 % de empresas registran solo deudas en moneda nacional (dolarización igual a 0), el 26.5 % solo deudas en moneda extranjera (dolarización igual a 1) y el 67.4 % deudas en ambas monedas (dolarización entre 0 y 1). El histograma del grado de dolarización de dichos deudores se muestra en la Figura 5.1.

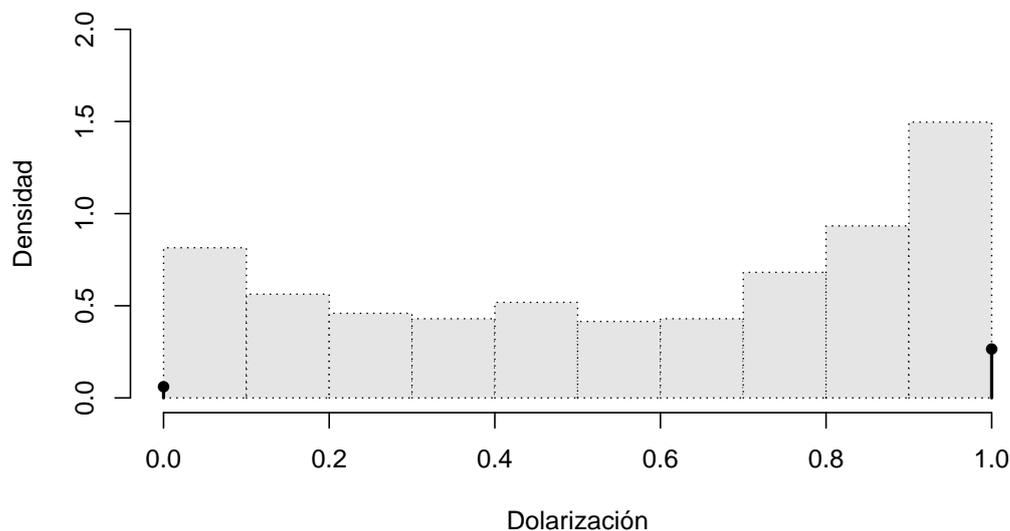


Figura 5.1: Histograma de la dolarización de créditos de deudores no minoristas

5. Las empresas registran información de deudas en el Sistema Financiero con un periodo mayor o igual a 6 meses previos a la fecha de corte de los estados financieros considerados.

5.4. Variables e indicadores

Para el análisis de variables y selección de indicadores, se consideró la información bajo la siguiente desagregación:

5.4.1. Información derivada de los estados financieros

Las variables de este bloque reflejan la situación financiera de la empresa, por lo que los indicadores seleccionados buscan reflejar distintos aspectos de un análisis del balance general y de los estados de resultados tales como los ratios de endeudamiento y de rendimiento. Sin embargo, para el entendimiento de los indicadores, es necesario establecer los siguientes conceptos:

- Pasivos corrientes: son deudas existentes que se espera sean satisfechas utilizando los activos corrientes de la empresa. Comprende las cuentas por pagar, los gastos acumulados, entre otros.

- Gastos financieros: son todas aquellas comisiones e intereses cobrados por los bancos o cualquier otra entidad por utilizar su dinero prestado o servicio de tipo financiero.
- Gastos de operación: son los gastos de ventas, gastos generales y administrativos.

En este sentido, los indicadores financieros considerados para el modelamiento son los señalados a continuación:

(a) Medidas de endeudamiento

Los ratios de endeudamiento miden el apalancamiento financiero de una empresa, es decir, el nivel de deudas que registra la empresa frente a los recursos con que cuenta. Al respecto, se considera la razón de gastos financieros ($\text{end_gasto_finan_pascte_me}$), que muestra la relación de los gastos financieros respecto a los pasivos corrientes. Se expresa como sigue:

$$\text{end_gasto_finan_pascte_me} = \frac{\text{Gastos financieros}}{\text{Pasivo corriente en ME}}$$

(b) Medidas de rendimiento

Los ratios de rendimiento miden la eficiencia de la empresa en el uso de los recursos. Al respecto, se considera la razón de gastos (ren_costo_total), que mide la capacidad de la gerencia para controlar los gastos. Se expresa como sigue:

$$\text{ren_costo_total} = \frac{\text{Costo de ventas} + \text{Gastos de operación} + \text{Gastos financieros}}{\text{Ventas}}$$

(c) Medidas de ingresos en moneda extranjera

Se considera el porcentaje de ingreso en ME (porc_ingreso_me), que corresponde al porcentaje de las ventas o ingresos generados por la actividad propia del negocio registradas en moneda extranjera, respecto al monto total de dichos conceptos.

5.4.2. Información derivada del Reporte Crediticio Consolidado

Las variables de este bloque reflejan la información de deuda y del comportamiento de pago de las empresas en el Sistema Financiero.

- (a) Dolarización de los créditos (y), es el porcentaje de créditos directos en moneda extranjera sobre el total de créditos directos. Para la conversión de los créditos en moneda extranjera, se utiliza el tipo de cambio contable del mes evaluado.
- (b) Número de meses que el saldo de deuda en el Sistema Financiero disminuyó en los últimos 6 meses ($\text{sf_veces_saldo_sf_disminuye}$).
- (c) Número de meses que el saldo de deuda en moneda extranjera en el Sistema Financiero aumentó en los últimos 6 meses ($\text{sf_veces_saldo_me_usd_aumen}$).

5.5. Resultados

A fin de evaluar el desempeño del modelo de regresión a la media simplex inflacionada definido en (3.1) para ajustar los datos de dolarización, se consideró conveniente realizar también el ajuste de los datos bajo el modelo de regresión beta inflacionada definido en (1.14). Para tal fin, considerando la función de enlace logística, se ajustó un conjunto de regresiones con cinco covariables para estimar los parámetros α_0 y α_1 y la media de la respuesta γ . Los resultados de las regresiones son detallados en la Tabla 5.1.

Parámetro	Covariable	Estimaciones	
		Simplex	Beta
α_0	(Intercept)	-0.4845 (0.6167)	-0.2493 (0.6111)
	porc_ingreso_me	-0.0033 (0.0048)	-0.0055 (0.0048)
	sf_veces_saldo_me_usd_aumen	-0.4191 (0.1405)	-0.4418 (0.1409)
	ren_costo_total	-0.6082 (0.3469)	-0.6577 (0.3386)
	end_gasto_finan_pascte_me	1.2171 (0.5177) **	1.1361 (0.4778) **
	sf_veces_saldo_sf_disminuye	0.0051 (0.1248)	-0.0090 (0.1254)
α_1	(Intercept)	-1.6422 (0.7219)	-1.5971 (0.7008)
	porc_ingreso_me	0.0175 (0.0039) **	0.0179 (0.0038) **
	sf_veces_saldo_me_usd_aumen	-0.1066 (0.1141)	-0.1294 (0.1129)
	ren_costo_total	-0.2326 (0.2556)	-0.2467 (0.2418)
	end_gasto_finan_pascte_me	-0.1086 (0.3976)	-0.1108 (0.4062)
	sf_veces_saldo_sf_disminuye	0.0980 (0.1177)	0.0962 (0.1186)
γ	(Intercept)	-2.4587 (0.3043)	-2.1802 (0.2912)
	porc_ingreso_me	0.0225 (0.0016) **	0.0219 (0.0016) **
	sf_veces_saldo_me_usd_aumen	0.2490 (0.0477) **	0.2402 (0.0482) **
	ren_costo_total	0.4723 (0.1656) **	0.4131 (0.1478) **
	end_gasto_finan_pascte_me	-0.9780 (0.2782)	-0.9387 (0.2474)
	sf_veces_saldo_sf_disminuye	0.2209 (0.0504) **	0.1688 (0.0513) **
σ^2	sigma	5.3762 (0.1790) **	2.2355 (0.1324) **

Cuadro 5.1: Coeficientes estimados para los modelos de regresión a la media simplex y beta inflacionadas. En paréntesis se muestra el error estándar de la estimación, obtenida de la inversa de la matriz de información de Fisher Observada evaluada en los parámetros estimados, y que permiten testear la significancia individual de los parámetros utilizando el estadístico de Wald (coeficientes significativos al 1% son señalados por **).

Al respecto, se observa que existe una relación positiva entre la media de la respuesta (grado de dolarización de las empresas) y las covariables, con excepción de la razón de gastos financieros. Sobre el particular, si bien esta variable resultó ser no significativa, se incluyó en el modelo dada su relación con la dolarización: el ratio se reduce cuando los pasivos corrientes en moneda extranjera se incrementan, por lo que las empresas buscarían opciones de financiamiento para afrontar dichas responsabilidades y con ello incrementar su grado de dolarización. Por otro lado, las demás variables resultaron significativas. En primer lugar, empresas que tienen sus ingresos principalmente en moneda extranjera, suelen tener sus obligaciones también en dicha moneda (mayor dolarización) para así calzar sus exposiciones netas. En segundo lugar, aquellas empresas que, en los últimos 6 meses, presentan incrementos continuos del saldo de deuda en moneda extranjera o reducciones del saldo de deuda total

en el Sistema Financiero, registran un mayor grado de dolarización por la propia naturaleza de cálculo del indicador. En tercer lugar, a menor valor de la razón de gastos (eficiencia en el uso de los recursos), la empresa puede afrontar los gastos con sus propios recursos y no requerir financiamientos alternativos (menor dolarización).

Finalmente, para evaluar el desempeño de los modelos, se ha analizado el criterio de información de akaike (AIC, por sus siglas en inglés) y el condicional (AICc), el criterio de información bayesiano (BIC, por sus siglas en inglés), el error cuadrático medio (MSE, por sus siglas en inglés) y el coeficiente de determinación R^2 , que son presentados en la Tabla 5.2. Al respecto, los resultados de todos los indicadores evidencian que el modelo de regresión a la media simplex inflacionada ajusta mejor los datos en comparación de la beta.

Criterio	Simplex	Beta
AIC	730.7	770.7
AICc	731.9	771.8
BIC	816.5	856.5
MSE	0.08902	0.08914
R^2	0.3321	0.3312

Cuadro 5.2: Criterios de comparación AIC, AICc, BIC, MSE y R^2 para los modelos de regresión a la media simplex y beta inflacionadas para el ajuste de los datos de dolarización.

Capítulo 6

Conclusiones

6.1. Conclusiones

Cuando se busca modelar variables independientes definidas en el intervalo abierto $(0, 1)$, como proporciones o probabilidades, es necesario considerar otras técnicas de estimación con el fin de obtener resultados confiables, tales como las distribuciones beta, beta rectangular y simplex.

Existen situaciones donde se requieren modelar probabilidades no nulas para los valores extremos 0 y 1. Una estrategia consiste en utilizar los modelos de regresión en dos etapas, sin embargo, la debilidad de estos modelos es que se torna más difícil la interpretación del efecto de las covariables sobre la media de la respuesta al depender de varios parámetros.

Ante este problema, el presente trabajo de tesis propone estudiar el modelo de regresión a la media simplex inflacionada, que permite modelar variables aleatorias continuas limitadas en el intervalo cerrado $[0, 1]$. La principal ventaja del modelo es que la estructura de regresión planteada permite analizar la influencia de las covariables directamente sobre la media, a diferencia de los modelos de regresión en dos etapas, que se realiza sobre la media condicional.

Mediante el estudio de simulación, se verificó que el método descrito permite recuperar los parámetros del modelo de regresión a la media simplex inflacionada y que los estimadores presentan menor sesgo relativo y error cuadrático medio y mayor cobertura a medida que aumenta el número de observaciones. Asimismo, mediante la aplicación del modelo para la estimación del grado de dolarización, se evidenció que el modelo de regresión simplex inflacionada ajusta mejor los datos en comparación con el modelo de regresión a la media beta inflacionada.

6.2. Sugerencias para investigaciones futuras

- Establecer criterios para la selección de variables a considerar para las ecuaciones de regresión para estimar la media de la respuesta y los parámetros que modelan las probabilidades de los valores extremos 0 y 1.
- Proponer variantes al modelo de regresión a la media simplex inflacionada que permita no linealidad y que el parámetro de dispersión σ^2 de los datos no sea constante para todas las observaciones, sino que dependa de un conjunto de regresores a través de una función de enlace.

Apéndice A

Resultados teóricos

En esta sección se detalla la función score y la matrix hessiana de la función log-verosimilitud (3.3).

La función score es dada por: $g(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial K(\boldsymbol{\theta})}{\partial \omega_j} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \eta_j} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \beta_j} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \sigma^2} \end{pmatrix}$,

donde, las derivadas de primer orden son:

$$\begin{aligned} \frac{\partial K(\boldsymbol{\theta})}{\partial \omega_j} &= \sum_{i=1}^{n_0} \frac{1}{\alpha_{0i} g'_1(\alpha_{0i})} \hat{x}_{ij} - \sum_{i=m+1}^n \frac{1-\gamma_i}{c_i g'_1(\alpha_{0i})} \hat{x}_{ij} \\ &\quad + \sum_{i=m+1}^n \left\{ \frac{(y_i - \mu_i)(\mu_i^2 - 2y_i \mu_i + y_i)}{\sigma^2 y_i (1-y_i) \mu_i^3 (1-\mu_i)^3} \right\} \frac{\gamma_i (1-\gamma_i)(1-\alpha_{1i})}{c_i^2 g'_1(\alpha_{0i})} \hat{x}_{ij} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \eta_j} &= \sum_{i=n_0+1}^m \frac{1}{\alpha_{1i} g'_2(\alpha_{1i})} \check{x}_{ij} - \sum_{i=m+1}^n \frac{\gamma_i}{c_i g'_2(\alpha_{1i})} \check{x}_{ij} \\ &\quad - \sum_{i=m+1}^n \left\{ \frac{(y_i - \mu_i)(\mu_i^2 - 2y_i \mu_i + y_i)}{\sigma^2 y_i (1-y_i) \mu_i^3 (1-\mu_i)^3} \right\} \frac{\gamma_i (1-\gamma_i)(1-\alpha_{0i})}{c_i^2 g'_2(\alpha_{1i})} \check{x}_{ij} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \beta_j} &= - \sum_{i=1}^{n_0} \frac{1}{(1-\gamma_i) g'_3(\gamma_i)} x_{ij} + \sum_{i=n_0+1}^m \frac{1}{\gamma_i g'_3(\gamma_i)} x_{ij} + \sum_{i=m+1}^n \frac{\alpha_{0i} - \alpha_{1i}}{c_i g'_3(\gamma_i)} x_{ij} \\ &\quad + \sum_{i=m+1}^n \left\{ \frac{(y_i - \mu_i)(\mu_i^2 - 2y_i \mu_i + y_i)}{\sigma^2 y_i (1-y_i) \mu_i^3 (1-\mu_i)^3} \right\} \frac{(1-\alpha_{0i})(1-\alpha_{1i})}{c_i^2 g'_3(\gamma_i)} x_{ij} \\ \frac{\partial K(\boldsymbol{\theta})}{\partial \sigma^2} &= \sum_{i=m+1}^n \frac{1}{\sigma^2} \left\{ \frac{1}{2\sigma^2 y_i (1-y_i)} \left(\frac{y_i - \mu_i}{\mu_i - \mu_i^2} \right)^2 - \frac{1}{2} \right\} \end{aligned}$$

La matriz Hessiana es dada por:

$$\begin{pmatrix} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \omega_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \eta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \beta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \sigma^2} \\ \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \omega_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \eta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \beta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \sigma^2} \\ \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \omega_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \eta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \beta_j} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \sigma^2} \\ \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \sigma^2 \partial \omega_h} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \sigma^2 \partial \eta_h} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \sigma^2 \partial \beta_h} & \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix},$$

donde, las derivadas de segundo orden son:

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \omega_j} &= - \sum_{i=1}^{n_0} \frac{g_1'(\alpha_{0i}) + \alpha_{0i} g_1''(\alpha_{0i})}{\alpha_{0i}^2 (g_1'(\alpha_{0i}))^3} \hat{x}_{ih} \hat{x}_{ij} + \sum_{i=m+1}^n \frac{(c_i g_1''(\alpha_{0i}) - (1 - \gamma_i) g_1'(\alpha_{0i})) (1 - \gamma_i)}{c_i^2 (g_1'(\alpha_{0i}))^3} \hat{x}_{ih} \hat{x}_{ij} \\ &+ \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i) (1 - \alpha_{1i})}{\sigma^2 y_i (1 - y_i)} \frac{\hat{h}_1(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i)}{(\mu_i^3 (1 - \mu_i)^3 c_i^2 g_1'(\alpha_{0i}))^2} \hat{x}_{ih} \hat{x}_{ij} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \eta_j} &= - \sum_{i=n_0+1}^m \frac{g_2'(\alpha_{1i}) + \alpha_{1i} g_2''(\alpha_{1i})}{\alpha_{1i}^2 (g_2'(\alpha_{1i}))^3} \check{x}_{ih} \check{x}_{ij} + \sum_{i=m+1}^n \frac{(c_i g_2''(\alpha_{1i}) - \gamma_i g_2'(\alpha_{1i})) \gamma_i}{c_i^2 (g_2'(\alpha_{1i}))^3} \check{x}_{ih} \check{x}_{ij} \\ &- \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i) (1 - \alpha_{0i})}{\sigma^2 y_i (1 - y_i)} \frac{\check{h}_2(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i)}{(\mu_i^3 (1 - \mu_i)^3 c_i^2 g_2'(\alpha_{1i}))^2} \check{x}_{ih} \check{x}_{ij} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \beta_j} &= - \sum_{i=1}^{n_0} \frac{(1 - \gamma_i) g_3''(\gamma_i) - g_3'(\gamma_i)}{(1 - \gamma_i)^2 (g_3'(\gamma_i))^3} x_{ih} x_{ij} - \sum_{i=n_0+1}^m \frac{g_3'(\gamma_i) + \gamma_i g_3''(\gamma_i)}{\gamma_i^2 (g_3'(\gamma_i))^3} x_{ih} x_{ij} \\ &- \sum_{i=m+1}^n \frac{(\alpha_{0i} - \alpha_{1i}) (c_i g_3''(\gamma_i) + (\alpha_{0i} - \alpha_{1i}) g_3'(\gamma_i))}{c_i^2 (g_3'(\gamma_i))^3} x_{ih} x_{ij} \\ &+ \sum_{i=m+1}^n \frac{(1 - \alpha_{1i}) (1 - \alpha_{0i})}{\sigma^2 y_i (1 - y_i)} \frac{\hat{h}_3(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i)}{(\mu_i^3 (1 - \mu_i)^3 c_i^2 g_2'(\alpha_{1i}))^2} x_{ih} x_{ij} \end{aligned}$$

$$\frac{\partial^2 K(\boldsymbol{\theta})}{\partial \sigma^2 \sigma^2} = \sum_{i=m+1}^n \frac{1}{(\sigma^2)^2} \left(\frac{-1}{\sigma^2 y_i (1 - y_i)} \left(\frac{y_i - \mu_i}{\mu_i - \mu_i^2} \right)^2 + \frac{1}{2} \right) x_{ih} x_{ij}$$

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \eta_j} &= - \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i)}{c_i^2 g_1'(\alpha_{0i}) g_2'(\alpha_{1i})} \hat{x}_{ih} \check{x}_{ij} \\ &+ \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i) \mu_i^2 (1 - \mu_i)^2}{\sigma^2 y_i (1 - y_i) g_1'(\alpha_{0i}) g_2'(\alpha_{1i})} \frac{A_1 - B_1}{(\mu_i^3 (1 - \mu_i)^3 c_i^2)} \hat{x}_{ih} \check{x}_{ij} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \beta_j} &= - \sum_{i=m+1}^n \frac{c_i + (1 - \gamma_i)(\alpha_{0i} - \alpha_{1i})}{c_i^2 g_1'(\alpha_{0i})^3 g_3'(\gamma_i)} \hat{x}_{ih} x_{ij} \\ &+ \sum_{i=m+1}^n \frac{(1 - \alpha_{1i}) \mu_i^2 (1 - \mu_i)^2}{\sigma^2 y_i (1 - y_i) g_1'(\alpha_{0i}) g_3'(\gamma_i)} \frac{A_2 + B_2}{(\mu_i^3 (1 - \mu_i)^3 c_i^2)^2} \hat{x}_{ih} x_{ij} \end{aligned}$$

$$\frac{\partial^2 K(\boldsymbol{\theta})}{\partial \omega_h \sigma^2} = - \sum_{i=m+1}^n \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) \gamma_i (1 - \gamma_i) (1 - \alpha_{1i})}{c_i^2 g_1'(\alpha_{0i}) (\sigma^2)^2 y_i (1 - y_i) \mu_i^3 (1 - \mu_i^3)} \hat{x}_{ih}$$

$$\begin{aligned} \frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \beta_j} &= - \sum_{i=m+1}^n \frac{c_i + (1 - \gamma_i)(\alpha_{0i} - \alpha_{1i})}{c_i^2 g_1'(\alpha_{0i})^3 g_3'(\gamma_i)} \hat{x}_{ih} x_{ij} \\ &+ \sum_{i=m+1}^n \frac{(1 - \alpha_{1i}) \mu_i^2 (1 - \mu_i)^2}{\sigma^2 y_i (1 - y_i) g_1'(\alpha_{0i}) g_3'(\gamma_i)} \frac{A_2 + B_2}{(\mu_i^3 (1 - \mu_i)^3 c_i^2)^2} \hat{x}_{ih} x_{ij} \end{aligned}$$

$$\frac{\partial^2 K(\boldsymbol{\theta})}{\partial \eta_h \sigma^2} = \sum_{i=m+1}^n \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) \gamma_i (1 - \gamma_i) (1 - \alpha_{0i})}{c_i^2 g_2'(\alpha_{1i}) (\sigma^2)^2 y_i (1 - y_i) \mu_i^3 (1 - \mu_i^3)} \hat{x}_{ih}$$

$$\frac{\partial^2 K(\boldsymbol{\theta})}{\partial \beta_h \sigma^2} = - \sum_{i=m+1}^n \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) (1 - \alpha_{1i}) (1 - \alpha_{0i})}{c_i^2 g_3'(\gamma_i) (\sigma^2)^2 y_i (1 - y_i) \mu_i^3 (1 - \mu_i^3)} x_{ih}$$

con las siguientes funciones:

$$\begin{aligned} \hbar_1(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i) &= \gamma_i (1 - \gamma_i) (1 - \alpha_{1i}) \mu_i^2 (1 - \mu_i)^2 [(6\mu_i y_i - 3\mu_i^2 - 2y_i^2 - y_i) \mu_i (1 - \mu_i) \\ &- 3(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)(1 - 2\mu_i)] \\ &- \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) \mu_i^3 (1 - \mu_i)^3 (c_i^2 g_1''(\alpha_{0i}) - 2c_i (1 - \gamma_i) g_1'(\alpha_{0i}))}{g_1'(\alpha_{0i})} \end{aligned}$$

$$\begin{aligned} \hbar_2(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i) &= - \gamma_i (1 - \gamma_i) (1 - \alpha_{0i}) \mu_i^2 (1 - \mu_i)^2 [(6\mu_i y_i - 3\mu_i^2 - 2y_i^2 - y_i) \mu_i (1 - \mu_i) \\ &- 3(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)(1 - 2\mu_i)] \\ &- \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) \mu_i^3 (1 - \mu_i)^3 (c_i^2 g_2''(\alpha_{1i}) - 2c_i (\gamma_i) g_2'(\alpha_{1i}))}{g_2'(\alpha_{1i})} \end{aligned}$$

$$\begin{aligned} \hbar_3(\gamma_i, \alpha_{0i}, \alpha_{1i}, \mu_i, c_i, y_i) = & (1 - \alpha_{1i})(1 - \alpha_{0i})\mu_i^2(1 - \mu_i)^2[(6\mu_i y_i - 3\mu_i^2 - 2y_i^2 - y_i)\mu_i(1 - \mu_i) \\ & - 3(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)(1 - 2\mu_i)] \\ & - \frac{(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)\mu_i^3(1 - \mu_i)^3 \left(c_i^2 g_3''(\gamma_i) + 2c_i g_3'(\gamma_i)(\alpha_{0i} - \alpha_{1i}) \right)}{g_3'(\gamma_i)} \end{aligned}$$

$$A_1 = \mu_i(1 - \mu_i)c_i(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)(2\gamma_i(1 - \alpha_{1i}) - c_i)$$

$$\begin{aligned} B_1 = & \gamma_i(1 - \gamma_i)(1 - \alpha_{0i}) \left\{ \mu_i(1 - \mu_i) \left[-2(1 - \alpha_{1i})(y_i - \mu_i)^2 \right. \right. \\ & \left. \left. - (\mu_i^2 - \mu_i y_i + y_i) + 3(1 - \alpha_{1i})(1 - 2\mu_i) \right] \right\} \end{aligned}$$

$$A_2 = \mu_i(1 - \mu_i)c_i(y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i) \left\{ (1 - 2\gamma_i)c_i - 2(\gamma_i - \gamma_i^2)(\alpha_{0i} - \alpha_{1i}) \right\}$$

$$\begin{aligned} B_2 = & (1 - \alpha_{0i})(1 - \alpha_{0i})c_i^{-2}\gamma_i(1 - \gamma_i)\mu_i(1 - \mu_i)c_i^2[(2y_i\mu_i - 2\mu_i^2 - 2y_i^2 + 2\mu_i y_i \\ & - \mu_i^2 + 2\mu_i y_i - y_i)\mu_i(1 - \mu_i)c_i^2 - (y_i - \mu_i)(\mu_i^2 - 2\mu_i y_i + y_i)3(1 - \mu_i c_i^2)] \end{aligned}$$

Derivadas consideradas:

$$\begin{aligned} \frac{\partial \alpha_{0i}}{\partial \omega_j} &= \frac{\hat{x}_{ij}}{g_1'(\alpha_{0i})} \\ \frac{\partial \alpha_{1i}}{\partial \eta_j} &= \frac{\check{x}_{ij}}{g_2'(\alpha_{1i})} \\ \frac{\partial \gamma_i}{\partial \beta_j} &= \frac{x_{ij}}{g_3'(\gamma_i)} \\ \frac{\partial \mu_i}{\partial \omega_j} &= \frac{\gamma_i(1 - \gamma_i)(1 - \alpha_{1i})}{c_i^2 g_1'(\alpha_{0i})} \hat{x}_{ij} \\ \frac{\partial \mu_i}{\partial \eta_j} &= \frac{-\gamma_i(1 - \gamma_i)(1 - \alpha_{0i})}{c_i^2 g_2'(\alpha_{1i})} \check{x}_{ij} \\ \frac{\partial \mu_i}{\partial \beta_j} &= \frac{(1 - \alpha_{0i})(1 - \alpha_{1i})}{c_i^2 g_3'(\gamma_i)} x_{ij} \\ \frac{\partial c_i}{\partial \beta_j} &= \frac{(\alpha_{0i} - \alpha_{1i})}{g_3'(\gamma_i)} x_{ij} \end{aligned}$$

Bibliografía

- Bandyopadhyay, D., Galvis, D. y Lachos, V. (2017). Augmented mixed models for clustered proportion data, *Statistical Methods in Medical Research* **26**(2): 880–897.
- Barndorff-Nielsen, O. y Jorgensen, B. (1991). Some parametric models on the simplex, *Journal of Multivariate Analysis* **39**(1): 106–116.
- Bayes, C., Bazan, J. y Garcia, C. (2012). A new robust regression model for proportions, *Bayesian Analysis* **7**(4): 841–866.
- Bayes, C. y Valdivieso, L. (2016). A beta inflated mean regression model for fractional response variables, *Journal of Applied Statistics* **43**(10): 1814–1830.
- Cox, C. (1996). Nonlinear quasi-likelihood models: applications to continuous proportions, *Computational Statistics & Data Analysis* **21**(4): 449–461.
- Cribari-Neto, F. y Zeileis, A. (2010). Beta regression in R, *Journal of Statistical Software* **34**(2): 1–24.
- Espinheira, P. y de Oliveira Silva, A. (2018). Nonlinear simplex regression models. eprint arXiv:1805.10843.
- Ferrari, S. y Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions, *Journal of Applied Statistics* **31**(7): 799–815.
- Galvis, D., Bandyopadhyay, D. y Lachos, V. (2014). Augmented mixed beta regression models for periodontal proportion data, *Statistics in Medicine* **33**(21): 3759–3771.
- Hahn, E. (2008). Mixture densities for project management activity times: A robust approach to PERT, *European Journal of Operational Research* **188**(2): 450–459.
- Jorgensen, B. (1997). *The theory of dispersion models*, Chapman & Hall.
- Kieschnick, R. y McCullough, B. (2003). Regression analysis of variates observed on (0,1): Percentages, proportions, and fractions, *Statistical Modelling* **3**(3): 193–213.
- Lopez, F. (2013). A bayesian approach to parameter estimation in simplex regression model: A comparison with beta regression, *Revista Colombiana de Estadística* **36**(1): 1–21.
- Maddala, G. (1991). A perspective on the use of limited-dependent and qualitative variables models in accounting research, *The Accounting Review* **66**(4): 788–807.
- Miyashiro, E. (2008). *Modelos de regressão beta e simplex para análise de proporções*, Tesis de maestría, Universidade de São Paulo.
- Ospina, R. y Ferrari, S. (2010). Inflated beta distributions, *Statistical Papers* **51**: 111–126.
- Papke, L. y Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *Journal of Applied Econometrics* **11**(6): 619–632.

- Ramalho, J. y da Silva, J. (2009). A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms, *Quantitative Finance* **9**(5): 621–636.
- Smithson, M. y Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables, *Psychological methods* **11**(1): 54–71.
- Song, P., Qiu, Z. y Tan, M. (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data, *Biometrical Journal* **46**(5): 540–553.
- Song, P. y Tan, M. (2000). Marginal models for longitudinal continuous proportional data, *Biometrics* **56**(2): 469–502.
- Superintendencia de Banca, S. y. A. (2005). *Reglamento para la Administración del Riesgo Cambiario Crediticio*. https://intranet2.sbs.gob.pe/dv_int_cn/1369/v1.0/Adjuntos/0041-2005.r.pdf.
- Superintendencia de Banca, S. y. A. (2008). *Reglamento para la Evaluación y Clasificación del Deudor y la Exigencia de Provisiones*. https://intranet2.sbs.gob.pe/dv_int_cn/1097/v4.0/Adjuntos/11356-2008.r.pdf.
- Superintendencia de Banca, S. y. A. (2011). *Reglamento de Gestión de Riesgo de Crédito*. https://intranet2.sbs.gob.pe/dv_int_cn/774/v3.0/Adjuntos/3780-2011.r.pdf.
- Zhang, P., Qiu, Z. y Shi, C. (2016). simplexreg: An R package for regression analysis of proportional data using the simplex distribution, *Journal of Statistical Software* **71**(11).