

Universidad Nacional de Ingeniería

FACULTAD DE INGENIERÍA ECONÓMICA, ESTADÍSTICA Y CIENCIAS

SOCIALES




TESIS

Regresión a la media para datos fraccionarios usando enlaces asimétricos con una aplicación a cobranzas de tarjetas de crédito

PARA OBTENER EL TÍTULO PROFESIONAL DE
INGENIERO ESTADÍSTICO

Elaborado por

Osman Dennys Rafael Taípe

 [0000-0001-7401-7428](https://orcid.org/0000-0001-7401-7428)

Asesor

Mag. José Bernardo Espinoza Suarez

 [0000-0002-0796-4975](https://orcid.org/0000-0002-0796-4975)

LIMA – PERÚ

2023

Citar/How to cite	Rafael Taipe,2023[1]
Referencias/Reference	N. Rafael Taipe. 2023, “Guía para la
Estilos/Style:	presentación de trabajos formales conducentes
IEEE	a la obtención de grados o títulos [Tesis de grado], Lima (Perú): Universidad nacional de Ingeniería 2023

Citar/How to cite	(Rafael Taipe,2023)
Referencias/Reference	(Rafael Taipe. 2023), guía para la presentación
Estilos/Style:	de trabajos formales conducentes a la
IEEE	obtención de grados o títulos [Tesis de grado, Universidad Nacional de Ingeniería]. Repositorio Institucional UNI.

Dedicatoria

A mi abuela Alejandrina,
mi esposa Vanessa,
mis padres Carmen y Ursisino,
hermanas Lorena y Katy
quienes me brindaron su apoyo incondicional.

Agradecimientos

A los profesores de la escuela de ingeniería estadística por transmitir su conocimiento, con un especial agradecimiento al profesor Hernan Belisario Garrafa Aragón.

A mi asesor y revisor de tesis por su paciencia y constante dedicación Jose Espinoza Suarez y Silva Guillen respectivamente.

A mis tíos Héctor Taipe Martínez, Fredy Taipe Martínez y Betty Taipe Martínez por su constante apoyo y consejos a lo largo de toda mi carrera, en especial a Héctor por ser como un padre y estar en los momentos difíciles.

Índice

Dedicatoria.....	II
Agradecimientos	III
Índice	IV
Índice de Tablas	VIII
índice de Gráficos	IX
Resumen.....	X
Abstract.....	XI
Capítulo I: Introducción.....	1
1.1 Generalidades	1
1.2 Planteamiento de la realidad problemática.....	3
1.3 Formulación del Problema	5
1.3.1 Problema general.....	5
1.3.2 Problema específico	5
1.4 Objetivo	5
1.4.1 Objetivos general.....	5
1.4.2 Objetivos específicos.....	5
1.5 Justificación.....	6
1.6 Hipótesis	6
1.6.1 Hipótesis General	6
1.6.2 Hipótesis Específicas.....	6
Capitulo II: Fundamento Teórico.....	7
2.1 Marco teórico modelos matemáticos y estadísticos	7
2.1.1 La distribución Beta Inflacionada	7
2.1.2 Una reparametrización de la distribución Beta Inflacionada	8

2.1.3 Distribuciones Exponenciales Potencia.....	10
2.1.4 Distribuciones de Valor Extremo Generalizado.....	10
2.1.5 El Teorema de Valores Extremos.....	12
2.1.6 Distribución de Gumbel Exponenciada.....	13
2.1.7 Regresión a la media para datos fraccionarios usando enlaces asimétricos.....	15
2.2 Marco Teórico relacionado al negocio.....	26
2.2.1 Riesgo de crédito.....	27
2.2.2 Riesgo país.....	28
2.2.3 Definiciones de estrategias de recupero en Banca.....	28
2.2.4 Modelos de scoring de cobranzas.....	29
2.2.5 La gestión de cobranzas en banca.....	31
2.2.6 Definición créditos perdida.....	32
Capítulo III: Metodología de la investigación.....	33
3.1 Tipo, Nivel y Diseño de la investigación.....	33
3.1.1 Tipo de investigación.....	33
3.1.2 Nivel de investigación.....	34
3.1.3 Diseño de investigación.....	34
3.2 Población de estudio.....	34
3.3 Diseño muestral.....	34
3.4 Técnicas e instrumento de recolección de datos.....	35
3.2 Operacionalización de variables.....	35
Capitulo IV: Procesamiento de los datos.....	39
4.1 Definición de variables cuantitativas.....	39
4.3 Conversión de variables cualitativas (nominales y ordinales) a Dummy:.....	42

4.4 Revisión de valores atípicos	42
4.5 Detección de atípicos univariados	42
4.6 Detección de atípicos multivariados.....	44
4.7 Aplicación del modelo	45
Capítulo V: Resultados	46
5.1 Resultados de ajuste de modelo	46
5.1.1 Resultados predictivos del modelo.....	50
Capítulo VI: Conclusiones.....	52
6.1 Aplicación del modelo a datos simulados	52
6.2 Aplicación del modelo a datos reales	53
Capitulo VII: Recomendaciones	56
Referencias Bibliográficas	57
Anexos	1
9.1 Cálculo Derivadas parciales y dobles.....	1
9.2 Tablas y gráficos de la simulación del modelo	8
9.3 Código en Matlab LOGLOG:	13
9.3.1 Función acumulada y sus derivadas	13
9.3.2 Simulación de distribuciones caso 0 inflacionado	13
9.3.4 Gradiente y Matriz Hessiana para 1 inflacionado	14
9.3.5 Gradiente y Hesiana para 0-1 Inflacionado	20
9.3.6 Simulacion de valores de la distribución y recuperaciond de parámetros caso 1 inflacionada	29
9.3.7 Simulacion de valores de la distribución y recuperaciond de parámetros caso 0-1 inflacionada	32
9.4 Código en Matlab CLOGLOG:.....	35

9.4.1 Función acumulada y sus derivadas	35
9.3.2 Simulación de distribuciones caso 1 inflacionado	35
9.3.3 Simulación de distribuciones caso 0-1 inflacionado	36
9.4.2 Gradiente y Matriz Hessiana para 1 inflacionado	37
9.4.3 Gradiente y Hesiana para 0-1 Inflacionado.....	41
9.4.4 Simulacion de valores de la distribución y recuperaciond de parámetros caso 0 inflacionada	50
9.4.5 Simulacion de valores de la distribución y recuperaciond de parámetros caso 0-1 inflacionada	53

Índice de Tablas

Tabla 1: Tabla de frecuencias para distribución de tres partes para la potencia $a=0.5$, $a=1.0$ y $a=1.5$	20
Tabla 2: Resumen de AIC para el modelo propuesto de enlaces asimétricos y logísticos para los 9 escenarios planteados en la simulación.....	21
Tabla 3: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico $n=1000$ caso0 inflacionado.....	22
Tabla 4: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico $n=1000$ caso1 inflacionado.....	23
Tabla 5: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico $n=1000$ caso 0-1 inflacionado	25
Tabla 6: Matriz de consistencia para el modelo de enlaces asimétricos	35
Tabla 7 : Operacionalización de las variables.....	37
Tabla 8: Presentación descriptiva de variables cuantitativas	39
Tabla 9: Presentación de variables cualitativas y ordinales	41
Tabla 10: Tabla de segmentos para identificación de outliers multivariados	44
Tabla 11: Comparación de resultados para metodología de enlaces asimétricos (loglog) y.....	48
enlaces logísticos para caso 0 inflacionada del mes1 y mes2	48
Tabla 12: Comparación de resultados para metodología de enlaces asimétricos (cloglog) y.....	49
enlaces logísticos para caso 0 inflacionada del mes2 y mes3	49
Tabla 13: Comparación de AIC, BIC para metodología de enlaces asimétricos y enlaces logísticos caso 0,1 y 0-1 inflacionado Mes1 a 3.....	50
Tabla 14: Tamaños de muestra para entrenamiento y validación del modelo	50
Tabla 15: Indicadores ECM y MAPE para la muestra de validación	51

índice de Gráficos

Figura 1 : Distribución acumulada Gumbel Estandar Potencia (a) vs Gumbel Estándar Potencia Complemento.....	15
Figura 2 : Distribución acumulada Gumbel Estandar Potencia (b) vs Gumbel Estándar Potencia Complemento.....	15
Figura 3 : Distribuciones simuladas usando enlaces CLOGLOG y LOGLOG potencia n=1000.....	20
Figura 4 Presentación de variables cualitativas y ordinales.....	40
Figura 5 Deteccion de valores atípicos univariados.....	43

Resumen

En esta investigación se propone un nuevo modelo de regresión a la media, para variables independiente fraccionarias usando enlaces asimétricos, estas variables toman valores continuos en el intervalo abierto (0,1) e inclusive tomando los valores extremos 0 y 1, estas variables podrían ser proporciones, tasas, ratios, etc.

Para el desarrollo de la tesis, tomamos como base la investigación desarrollada por (Bayes & Valdivieso, 2016), la cual ajusta el conjunto de ecuaciones para el rango (0,1) y los extremos 0 y 1 en un solo proceso de estimación a la media, por lo cual puede ser trabajado con funciones de enlace arbitrarios. Para este caso especial (Bayes & Valdivieso, 2016) por conveniencia usaron un enlace de tipo logístico $g(\gamma) = x'\beta$, donde g es una función de enlace y x es un vector de covariables.

Los enlaces propuestos son de tipo loglog y cloglog que se derivan de la distribución Gumbell Estandar Complementaria Exponenciada, el cual fue trabajado por (Gupta & Kundu, 1999) bajo el supuesto de un mejor ajuste a los datos. La optimización de la función log-verosimilitud se desarrolló mediante un algoritmo de puntuación de Fisher modificando utilizado una rutina de matlab con el vector de gradiente analítico y la matriz Hesiana.

Para probar las hipótesis, se desarrolló un análisis de simulación comparativo, con el objetivo de recuperar los parámetros y comparar el ECM para el modelo propuesto vs el modelo de (Bayes & Valdivieso, 2016). Adicionalmente se aplicó el modelo desarrollado a datos de cobranzas de tarjetas de crédito para un banco peruano.

Palabras-clave: Variables dependientes fraccionarias, Distribucion Beta Inflacionada, Distribucion Gumbell Exponenciada, enlaces loglog, enlaces cloglog.

Abstract

In this research a new regression model to the mean is proposed, for fractional independent variables using asymmetric links, these variables take continuous values in the open interval (0,1) and even taking the extreme values 0 and 1, these variables could be proportions, rates, ratios, etc.

For the development of the thesis, we take as a basis the research developed by (Bayes & Valdivieso, 2016) that adjusts the set of equations for the range (0,1) and the extremes 0 and 1 in a single process of estimating the mean $g(\gamma) = x'\beta$, so it can be worked with arbitrary link functions. For this special case (Bayes & Valdivieso, 2016) for convenience they used a logistic type link, where g is a link function and x is a vector of covariates.

The proposed links are of type Loglog and Cloglog derived from the Gumbell Standard Complementary Exponentiated distribution, which was worked on by (Gupta & Kundu, 1999) under the assumption of a better fit to the data. The optimization of the log-likelihood function was developed using a Fisher scoring algorithm modifying using a matlab routine with the analytical gradient vector and the Hessian matrix.

To test the hypotheses, a comparative simulation analysis was developed, with the objective of recovering the parameters and comparing the ECM for the comparison of the proposed model vs the model of (Bayes & Valdivieso, 2016). Additionally, the developed model was applied to data on credit card collections for a Peruvian bank.

Keywords: Fractional Dependent Variables, Influenced Beta Distribution, Exponential Gumbel Distribution, loglog links, cloglog links.

Capítulo I: Introducción

1.1 Generalidades

Múltiples investigaciones de diferentes campos están interesadas en estudiar y medir la influencia de un conjunto de covariables sobre una proporción. Existe una gran cantidad de regresiones de modelos para proporciones en la literatura, por ejemplo: (Cribari-Neto & Ferrari, 2004) proponen un modelo de regresión beta basado en una reparametrización de la distribución beta. Por otro lado (Bayes, Bazán, & García, 2012) proponen un modelo de regresión rectangular beta que es una mezcla de una beta y una distribución uniforme y resistente a la presencia de valores atípicos.

Aunque estos modelos son adecuados y útiles para el análisis de regresión de proporciones, suponen que estas proporciones están restringidas al intervalo abierto continuo en el rango (0,1). Sin embargo, las variables podrían tomar los valores 0 o 1 en datos reales, dando lugar a las llamadas variables acotadas. Las siguientes investigaciones muestran modelos que ingresan los valores acotados de la variable:

- Modelos basados en transformaciones: la variable dependiente se transforma a una variable acotada fraccional en un intervalo unitario cerrado a un intervalo unitario abierto, y luego cualquier modelo de regresión para variables limitadas como beta o beta rectangular pueden ser aplicados. (Smithson & Verkuilen, 2006) proponen, por ejemplo, para transformar una variable acotada fraccional de la siguiente manera:

$$Y^* = \frac{Y(N - 1) + 0.5}{N}$$

- Modelos de dos partes: Primero, se aplica un modelo de regresión multinomial a una variable categórica que clasifica la respuesta fraccionaria en 0, 1 o cualquier valor en el intervalo continuo abierto (0, 1). Luego, condicional al hecho de que la variable respuesta

se encuentre en el intervalo (0,1), se aplica cualquier modelo de regresión para variables acotadas, como la regresión beta. En la investigación de (Ramalho & da Silva, 2009) se propone un modelo de respuesta fraccional, donde la probabilidad que la variable de respuesta toma el valor 0 o 1 se estima mediante un modelo binario, y luego se establece otro modelo cuando el valor de la variable se encuentra en un intervalo continuo abierto (0,1). Por otro lado, (Ospina & Ferrari, 2010) proponen diferentes modelos para intervalos (0,1] (beta inflado en uno, denotado por BEOI), [0; 1) (beta inflado en cero, denotada por BEZI) y [0;1] (beta inflado en 0-1, denotado por BEINF). Para los modelos BEOI y BEZI, se propone una mezcla de Distribución Bernoulli y se propone una distribución beta para BEINF una mezcla de Multinomial y beta. Vale la pena señalar que el valor esperado de estas mezclas está compuesto por diferentes parámetros, por lo que los efectos sobre la media son difíciles de interpretar bajo estos modelos.

- Modelos de una parte: en este tipo de modelos, la respuesta media se modela directamente con el conjunto de covariables. En (Papke & Wooldridge, 1996), está directamente relacionado con el vector de covariables a través de la ecuación $g(\gamma) = x'\beta$, donde g es la función de enlace. La estimación de este modelo se basa en una metodología de cuasi-verosimilitud que maximiza una función logarítmica de verosimilitud de Bernoulli, lo que lleva a una probabilidad cuasi máxima estimador (QMLE). Una propiedad de QMLE es el hecho de que es consistente y asintóticamente normal e independientemente de la distribución de la variable de respuesta condicional a covariables.

Por otro lado, (Bayes & Valdivieso, 2016) proponen una beta inflada modelo de regresión media que, basado en una conveniente reparametrización, permite modelar

directamente la media de la variable fraccional acotada de interés; estimación de este modelo se lleva a cabo desde una perspectiva clásica.

1.2 Planteamiento de la realidad problemática

En diversas situaciones es posible encontrar variables dependientes continuas que asumen valores en el intervalo de $(0,1)$ y toman los valores extremos 0 y 1. Estos valores podrían ser proporciones, tasas, ratios, etc. Claramente usar modelos de regresión lineal clásicos para explicar a la variable dependiente con la ayuda de covariables es inadecuado, ya que podrían producir valores ajustados para la variable de interés que exceda el límite superior o inferior del intervalo $[0,1]$. Otra alternativa de solución es transformar la variable respuesta de forma que asume valores en la línea recta y posteriormente modelar la media de la variable respuesta transformada. Este método sin embargo tiene inconvenientes al momento de interpretar los parámetros del modelo, la investigación de (Smithson & Verkuilen, 2006) muestra claramente el proceso de transformación de la variable. Otro problema es que generalmente los datos analizados pueden presentar asimetría ya que los valores de los extremos 0 y 1 están concentrados en las colas y por lo tanto la inferencia bajo el supuesto de normalidad puede no ser la adecuada. Los datos al presentar colas pesadas 0 o 1 inflacionados tienen problemas al ser ajustados por modelos de regresión a la media, que utilizan enlaces logísticos, una aplicación de lo expuesto se puede revisar en (Bayes & Valdivieso, 2016)

Desde una aplicación práctica, el uso de la regresión a la media con enlaces asimétricos se encuentra contextualizada al área de cobranzas de un banco del mercado peruano. El recupero de un cliente asociado a un producto, es un ratio que va de $(0,1)$ incluyendo los extremos, se mide con la expresión A/B siendo A el monto de pago del cliente y B la deuda total del cliente en un periodo de tiempo. En este caso analizaremos la recuperación de la deuda que tiene una

tarjeta de crédito de tipo Z; es una tarjeta nueva con beneficios poco conocidos y con poca participación en el mercado por lo cual el nivel de caer en default dentro del primer mes es alto. Si el cliente no paga su deuda en el rango de 0 a 30 días su perfil de riesgo será de problemas potenciales, dentro de 31 a 60 días será deficiente, 61 a 120 días será dudoso y 120 a más será de pérdida total de deuda.

Un problema que actualmente tiene el área de cobranzas del banco peruano es estimar el Ratio de Recupero (denominada RR) de los futuros clientes que caerán en default correspondientes al producto Z, esto con la ayuda de covariables del cliente. Y en base a los resultados de la estimación, realizar una gestión de cobranza más reciente de la cartera de clientes de la tarjeta Z.

Por lo expuesto anteriormente, se propone extender el desarrollo del modelo beta inflacionada desarrollado por (Bayes & Valdivieso, 2016) modificando los enlaces logísticos utilizados mediante la distribución Gumbel Potencia y Gumbel Potencia Complemento, ellos generan enlaces asimétricos de tipo cloglog potencia. Investigaciones como la de (Calabrese, 2012) proponen precisamente utilizar la distribución de Gumbel y más generalmente la de Valor Extremo Generalizado (GEV) para generar enlaces basados en variables dependientes binarias. En este trabajo buscamos usar estas ideas con la distribución de Gumbel Potencia, pero para variables inflacionadas; aunque a diferencia del enfoque bayesiano usado por (Wang & Dipak, 2010) usaremos un enfoque basado en la metodología de máxima verosimilitud.

Cabe resaltar que la distribución Gumbel Exponenciada que desarrollaremos en este trabajo está en base al trabajo de (Nadarajah, 2005).

1.3 Formulación del Problema

1.3.1 Problema general

¿Es posible desarrollar un modelo de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia y desarrollar un análisis comparativo con el modelo planteado por (Bayes & Valdivieso, 2016)?

1.3.2 Problema específico

¿Es posible realizar un análisis comparativo de desempeño usando datos de cobranzas de tarjetas de crédito, aplicando el modelo propuesto de enlaces asimétricos y el modelo desarrollado por (Bayes & Valdivieso, 2016)?

¿Es posible identificar y cuantificar cuáles son las variables que tienen más influencia en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito?

1.4 Objetivo

1.4.1 Objetivos general

Proponer un modelo de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia y realizar un análisis comparativo con el modelo desarrollado por (Bayes & Valdivieso, 2016)

1.4.2 Objetivos específicos

Realizar un análisis comparativo usando datos de cobranzas a tarjetas de crédito, aplicando el modelo propuesto de enlaces asimétricos y el modelo desarrollado por (Bayes & Valdivieso, 2016).

Identificar y cuantificar cuáles son las variables que tienen más influencia en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito.

1.5 Justificación

La presente investigación, se justifica de manera teórica en su aporte estadístico al marco de modelos de regresión a la media en variables independientes fraccionarias, usando enlaces asimétricos; proponiendo un modelo novedoso en base al enlace, tomando la distribución Gumbel Exponenciada y su complemento, el cual hace que el ajuste de los datos sea mejor al aplicarlos sobre distribuciones con colas pesadas de tipo 0 y 1 inflacionadas.

Otro aporte práctico se relaciona a la aplicación del modelo a distintos campos donde se pueda predecir y/o analizar variables dependientes de tipo fraccionaria, tal es el caso de la investigación, donde aplicamos el modelo propuesto a empresas bancarias donde se quiere predecir el ratio de recupero en el área de cobranzas.

1.6 Hipótesis

1.6.1 Hipótesis General

El modelo propuesto de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia tiene un menor ECM que el modelo desarrollado por (Bayes & Valdivieso, 2016) bajo el análisis de simulación comparativo.

1.6.2 Hipótesis Específicas

El modelo propuesto de enlaces asimétricos tiene un menor ECM que el modelo desarrollado por (Bayes & Valdivieso, 2016) usando los datos de cobranzas de una tarjeta de crédito.

El peso de la importancia de las variables demográficas es mayor al de las variables de negocio en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito.

Capítulo II: Fundamento Teórico

2.1 Marco teórico modelos matemáticos y estadísticos

En esta sección vamos a citar y cuestionar a todos los modelos matemáticos y estadísticos mencionados anteriormente y los cuales vamos a usar en la investigación de regresión a la media para datos fraccionarios usando enlaces asimétricos.

2.1.1 La distribución Beta Inflacionada

Una variable aleatoria Y tiene una distribución beta inflacionada, si su función de densidad de probabilidad es la siguiente: $Y \sim BI(\delta_0, \delta_1, \mu, \varphi)$

$$f(Y/\delta_0, \delta_1, \mu, \varphi) = \begin{cases} \delta_0 & ,si \quad Y=0 \\ (1-\delta_0-\delta_1)b(Y/\mu, \varphi) & ,si \quad Y \in (0,1) \\ \delta_1 & ,si \quad Y=1 \end{cases}$$

Donde $P(Y=0) = \delta_0$, $P(Y=1) = \delta_1$, $E(Y/Y \in (0,1)) = \mu$ φ es interpretado como el parámetro de precisión. La expresión $b(Y/\mu, \varphi)$ es la función de densidad de probabilidad de la distribución Beta, donde su media y varianza están representadas por μ y $\frac{\mu(1-\mu)}{1+\varphi}$

La media y la varianza de la distribución Beta Inflacionada viene dada por:

$$E(Y) = \delta_1 + (1 - \delta_0 - \delta_1)\mu$$

$$Var(Y) = \delta_1(1 - \delta_1) + (1 - \delta_0 - \delta_1)\left[\frac{\mu(1 - \mu)}{1 + \varphi} + (\delta_0 + \delta_1)\mu^2 - 2\mu\delta_1\right]$$

2.1.2 Una reparametrización de la distribución Beta Inflacionada

Con el fin de obtener una estructura de regresión simple para la media de Y , y permitir que los parámetros se puedan estimar sin ningún tipo de restricción, se propone usar la reparametrización alternativa planteada por (Bayes & Valdivieso, 2016) y es como sigue:

$$\gamma = \delta_1 + (1 - \delta_0 - \delta_1)\mu, \quad \alpha_0 = \frac{\delta_0}{1 - \gamma}, \quad \alpha_1 = \frac{\delta_1}{\gamma}$$

$$\text{donde } 0 \leq \gamma \leq 1, \quad 0 \leq \delta_0 \leq 1, \quad 0 \leq \delta_1 \leq 1$$

Bajo la nueva reparametrización se obtiene la siguiente función de densidad de probabilidad de la distribución Beta Inflacionada, de parámetros $\alpha_0, \alpha_1, \varphi, \gamma$. En adelante usaremos para la distribución de una variable aleatoria de ese tipo de notación:

$$Y \sim \text{BIm}(\alpha_0, \alpha_1, \gamma, \varphi)$$

$$f(Y / \alpha_0, \alpha_1, \gamma, \varphi) = \begin{cases} \alpha_0(1 - \gamma) & , \text{si } Y = 0 \\ \alpha(1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)b\left(Y / \frac{\gamma(1 - \alpha_1)}{1 - \alpha_0(1 - \gamma) - \alpha_1\gamma}, \varphi\right) & , \text{si } Y \in (0, 1) \\ \alpha_1\gamma & , \text{si } Y = 1 \end{cases}$$

La media y la distribución son dadas respectivamente por:

$$\gamma = E(Y), \quad \text{Var}(Y) = \left(\frac{1 + \alpha_1}{1 + \gamma}\right)\gamma + \left(\frac{(1 - \alpha_1)^2\varphi}{(1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)(1 + \varphi)} - 1\right)\gamma^2$$

Para los casos donde $\delta_0 = 0$ y $\delta_1 > 0$ se tiene una distribución inflacionada en unos, si $\delta_0 > 0$ se tiene una distribución inflacionada en ceros y si $\delta_0 = 0$ y $\delta_1 = 0$ se tiene distribución

beta inflacionada. Para el primer caso se tiene $\gamma = \delta_1 + (1 - \delta_0 - \delta_1)\mu$ y $\alpha_1 = \frac{\gamma_1}{\gamma}$ para el segundo caso $\gamma = (1 - \delta_0)\mu$ y $\alpha_1 = \frac{\gamma_0}{1-\gamma}$

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 proceden a ajustar un conjunto de ecuaciones de regresión a la media para los parámetros $\theta = (\omega, \eta, \beta, \phi)$ y su distribución para cada observación es la siguiente:

$$Y_i \sim BIm(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi)$$

$$f(Y_i / \alpha_{0i}, \alpha_{1i}, \gamma_i, \phi) = \begin{cases} \alpha_{0i}(1 - \gamma_i) & , si \ y = 0 \\ c_i b(Y_i / \frac{\gamma_i(1 - \alpha_{1i})}{1 - \alpha_{0i}(1 - \gamma_i) - \alpha_{1i}\gamma_i}, \phi) & , si \ y \in (0, 1) \\ \alpha_{1i}\gamma_i & , si \ y = 1 \end{cases}$$

$$\alpha_{0i} = g_{a1}^{-1}(\tilde{x}_i^t \omega), \quad \alpha_{1i} = g_{a2}^{-1}(\tilde{x}_i^t n), \quad \gamma_i = g_{a3}^{-1}(x_i^t \beta),$$

$$c_i = 1 - \alpha_{0i}(1 - \gamma_i) - \alpha_{1i}\gamma_i, \quad \mu_i = \frac{\gamma_i(1 - \alpha_{1i})}{c_i}$$

Donde $\omega = (\omega_0, \omega_1, \omega_2, \dots, \omega_{k1})$, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{k2})$, $n = (n_0, n_1, n_2, \dots, n_{k3})$ son vectores columna con los parámetros de la regresión $\tilde{x}_i = [1, x_i, \dots, x_{ik1}]^t$, $\check{x}_i = [1, x_i, \dots, x_{ik2}]^t$, $x_i = [1, x_i, \dots, x_{ik3}]^t$ son vectores columna de las covariables por individuo i , y g_{a1} , g_{a2} , g_{a3} son funciones de enlace estrictamente monótonas y diferenciales de segundo orden de $(0, 1) \mathbb{R}$; bajo este enfoque se puede analizar la influencia de las covariables directamente sobre la media, a diferencia de los modelos de dos partes que se realizaban sobre la media condicional.

2.1.3 Distribuciones Exponenciales Potencia

En la literatura (Gupta & Kundu, 1999) desarrolla una generalización para las distribuciones exponenciadas o potencia de dos maneras, si $G(x, \theta)$ es la función de distribución acumulada de una distribución base, con vector de parámetros θ , ellos añadieron un parámetro de potencia adicional α , para generar la función de distribución acumulada $F(x, \theta, \alpha)$ definida por:

Para el caso Máximo:

$$F(x, \theta, \alpha) = [G(x, \theta)]^\alpha$$

Para el caso Mínimo:

$$F(x, \theta, \alpha) = 1 - [1 - G(x, \theta)]^\alpha$$

Para $\alpha > 0$ es también llamado como el parámetro de potencia

2.1.4 Distribuciones de Valor Extremo Generalizado

Los valores extremos han constituido desde hace mucho tiempo una disciplina de gran interés, y no sólo para estadísticos sino también, para científicos e ingenieros. Existen varias definiciones en la literatura sobre este tema, algunas definiciones de estos autores se muestran a continuación.

“La teoría de valores extremos es una disciplina que desarrolla técnicas y modelos para describir los sucesos menos comunes, lo cual, para él, hace que sea una disciplina única. En cambio, para Gumbel (1958), el objetivo de la teoría de valores extremos es analizar valores extremos observados y predecir valores extremos en el futuro”.(Coles, 2001)

“Una definición más simplista que mencionan algunos autores es decir que los valores extremos son el máximo y el mínimo. La interpretación de lo que es algo extremo es complicada ya que su definición engloba varios atributos tales como excepcional, sorprendente y

catastrófico. Según dichos autores, al ser como se ha dicho subjetivamente difícil definir a los valores extremos, es mejor caracterizarlos mediante, por ejemplo, sus propiedades estadísticas, observaciones, predictibilidad, mecanismos, etc.” (Albeverio, Jentsch, & Kantz, 2005)

“Respecto a la antigüedad de esta teoría, afirman que puede decirse que tiene alrededor de 80 años de antigüedad, aunque el origen la teoría de valores extremos es bastante más antigua. Por su parte, Coles afirma que no fue hasta 1950 cuando se propuso una metodología seria para modelizar sucesos de este tipo. Asimismo, también dice que las primeras aplicaciones fueron en el campo de la ingeniería civil, pues según él, los ingenieros siempre han necesitado diseñar sus estructuras de forma que éstas soportaran las distintas fuerzas que podrían afectarlas (refiriéndose principalmente a fuerzas de la naturaleza)”. (Leadbetter, Lindgren, & Rootzén, 1983).

“Es también interesante comentar que la teoría asintótica de valores extremos se ha estudiado paralelamente a la del Teorema Central del Límite, de ahí que ambas teorías tengan bastante semejanza. Para ambos autores, la teoría de extremos está relacionada con el comportamiento límite del $\max\{x_1, x_2, x_3, \dots, x_n\}$ o del $\min\{x_1, x_2, x_3, \dots, x_n\}$ o cuando n tiende al infinito”. (Haan & Ferreira, 2006)

Algunos autores señalan que la teoría de valores extremos tuvo como precursor a Leonard Tippett, empleado de la British Cotton Industry Research Association, donde trabajaba para construir hilos de algodón más fuertes. En sus estudios, durante los años 20, se dio cuenta de que la fuerza de un hilo dependía de la fuerza de sus fibras más débiles. Pero en cambio, según Gumbel, el pionero a la hora de estudiar los valores más grandes para otras distribuciones fue E. L. Dodd, en 1923, aunque pocos siglos antes ya se habían empezado a estudiar los extremos a través de la distribución Normal.

2.1.5 El Teorema de Valores Extremos

En este capítulo del trabajo se va a ver la distribución de Valores Extremos Generalizada (cuyas siglas son GEV), que también es conocida como la distribución de Fisher-Tippett, la distribución tipo Von Mises-Jenkinson o la distribución de valores extremos tipo Von Mises. Según (Kotz & Nadarajah, 2001), dicha distribución fue inicialmente introducida por Jenkinson (1955).

Sean $x_1, x_2, x_3 \dots x_n$ variables aleatorias, $M_n = \text{Max}\{x_1, x_2, x_3 \dots x_n\}$ el máximo de dichas variables. Según Coles, en ocasiones, en la práctica las X_i (que tienen una función de distribución común F) corresponden a valores de un proceso medido en una escala regular de tiempo como medidas de niveles del mar cada hora, o temperaturas medias diarias de forma que representa M_n el máximo del proceso sobre n unidades temporales de observación. Si por ejemplo, n es el número de observaciones en un mes, entonces M_n corresponde al máximo mensual.

La función de distribución de M_n , que es como la gran mayoría de autores denotan al máximo de $x_1, x_2, x_3 \dots x_n$ se dedujo en su trabajo sobre estadísticos de orden, y es:

$$\begin{aligned} P[M_n \leq z] &= P[X_1 \leq z, X_2 \leq z, \dots, X_n \leq z] = P[X_1 \leq z]P[X_2 \leq z] \dots P[X_n \leq z] \\ &= F(z)F(z) \dots F(z) = (F(z))^n \end{aligned}$$

Según de (Haan & Ferreira, 2006), dicha función converge en probabilidad a 0 si $z < z^*$ y a 1 si $z \geq z^*$ donde $z^* = \sup\{x: F(x) < 1\}$ por tanto, que la distribución limite no sea degenerada, hay que tipificar, esto es, encontrar sucesiones de constantes $a_n > 0$ y de $b_n > 0$ ($n = 1, 2, \dots$) forma que la expresión $\frac{M_n - b_n}{a_n}$. Tenga una distribución no degenerada cuando $n \rightarrow \infty$ esto es que: $\lim_{n \rightarrow \infty} F^n(a_n + z_n) = G(z)$. Se trata de “estandarizar” la variable mediante una transformación lineal. A la nueva variable estandarizada se le llama M_n . Así, se tiene el

siguiente teorema, conocido como Teorema de Valores Extremos o Teorema de Fisher-Tippett-Gnedenko:

Teorema 1:

Si existen sucesiones de contrastes $\{a_n > 0\}$ y $\{b_n\}$ de forma que $P[\frac{M_n - b_n}{a_n} \leq Z] \rightarrow G(Z)$.

Cuando $n \rightarrow \infty$, donde G es la función de distribución no degenerada, entonces G pertenece a una de las siguientes familias:

$$I: G(z) = \exp\{-\exp[-(\frac{z-a}{b})]\}, \quad -\infty < z < \infty;$$

$$II: G(z) = \exp[-(\frac{z-a}{b})^{-\alpha}], \quad z > b, \quad z \leq b;$$

$$III: G(z) = \exp(-[-(\frac{z-a}{b})^{-\alpha}]), \quad z \leq b, \quad z > b;$$

Para parámetros $a_n > 0$, b y, en caso de las familias II y III donde $\alpha > 0$. Estas tres clases de distribuciones son conocidas como las distribuciones de valores extremos, donde las de tipo I son la familia de Gumbel; las de tipo II la de Fréchet, y las de tipo III la de Weibull, cada una con su parámetro de localización b y de escala a y además, las familias de Fréchet y de Weibull tienen un parámetro de forma α Más adelante se estudiarán las características de esas distribuciones, y se profundizará más en ellas.

2.1.6 Distribución de Gumbel Exponenciada

El trabajo de (Nadarajah, 2005) desarrolló una generalización de la distribución Gumbel en aplicaciones del modelado del clima. La distribución Gumbel se conoce también como la distribución de valores extremos de TIPO I. Algunas de sus áreas de aplicación en modelización incluyen:

Patrones de lluvia, calentamiento global, modelos de velocidad del viento, etc. En este apartado se propone una generalización de la distribución Gumbel la cual tiene mayor aplicabilidad a más problemas. Y sobre todo no presenta restricciones en los parámetros como si lo hacen las otras distribuciones de valor extremo. La generalización es motivada por el trabajo de (Gupta & Kundu, 1999) donde la distribución exponencial exponenciada (EE) es introducida como generalización de la distribución exponencial estándar. En particular la distribución EE se define por la función de distribución acumulada: $F(x; \lambda; \alpha) = (1 - \exp(-\lambda x))^\alpha, \alpha > 0$

Las propiedades matemáticas de esta distribución EE son estudiadas en detalle por (Gupta & Kundu, 1999) En este artículo nosotros introducimos una distribución que generaliza la distribución Gumbel estandar en la misma manera que se generalización la distribución EE. Dadas las distribuciones Gumbel Potencia y Gumbel Potencia Complementaria.

Distribución Gumbel Potencia

$$F(x; \mu; \theta) = [\exp(-\exp(-\frac{x-\mu}{\sigma}))]^\alpha$$

Distribución Gumbel Potencia Complemento

$$F(x; \mu; \theta) = 1 - [\exp(-\exp(\frac{x-\mu}{\sigma}))]^\alpha$$

De manera más particular dado que todos los parámetros μ y σ no podrán ser identificables en un contexto de estimación nosotros trabajamos con sus versiones estándar las cuales se dan cuando $\mu = 0$ y $\sigma = 0$. Más explícitamente ellas vienen dadas por:

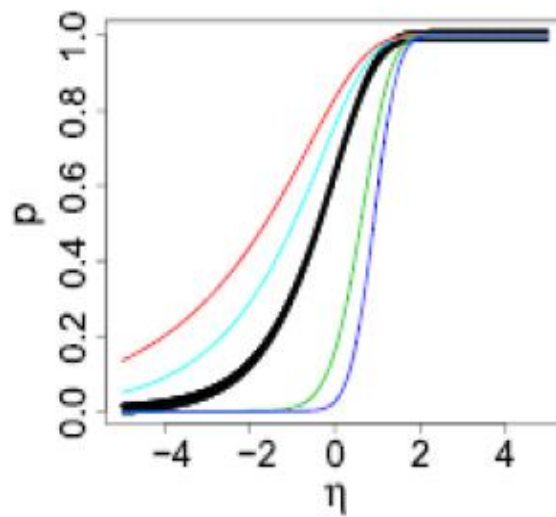
Distribución Gumbel Estandar Potencia

$$F(x; \mu; \theta) = [\exp(-\exp(-x))]^\alpha$$

Distribución Gumbel Estandar Potencia Complemento

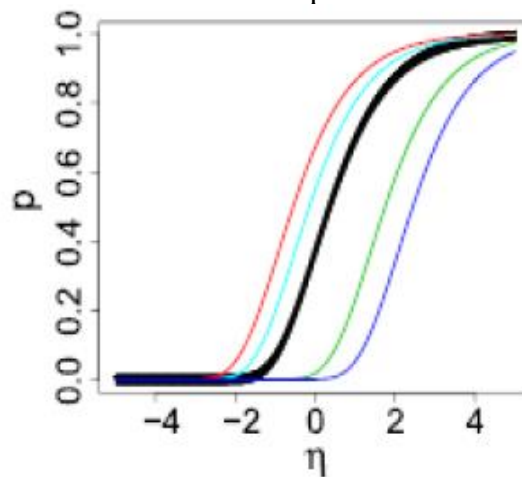
$$F(x; \mu; \theta) = 1 - [\exp(-\exp(x))]^\alpha$$

Figura 1 : Distribución acumulada Gumbel Estandar Potencia



Nota: $\alpha=0.4$ (rojo), $\alpha=0.6$ (celeste), $\alpha=1$ (negro), $\alpha=4$ (verde), $\alpha=8$ (azul)

Figura 2 : Distribución Gumbel Estandar Potencia Complemento



Nota: $\alpha=0.4$ (rojo), $\alpha=0.6$ (celeste), $\alpha=1$ (negro), $\alpha=4$ (verde), $\alpha=8$ (azul)

2.1.7 Regresión a la media para datos fraccionarios usando enlaces asimétricos

2.1.7.1 Funciones de enlace para Distribución Gumbel Exponenciada

En esta sección se desarrolla las funciones de enlace y su relación con las funciones de distribución. De la sección 2.1.6 donde $a=1/\alpha$ se desprenden las siguientes funciones de enlace de forma genérica:

Enlace loglog exponenciada: Gumbel Estandar Exponenciada

$$g(\mu_i) = -\log\left(-\frac{1}{a}\log(\mu_i)\right)$$

Enlace cloglog exponenciada: Gumbel Estandar Exponenciada Complemento

$$g(\mu_i) = \log\left(-\frac{1}{a}\log(1 - \mu_i)\right)$$

2.1.7.2 Definición del modelo

Planteamos el modelo de variable dependiente fraccionaria, representada por la función de densidad de probabilidad $Y_i \sim BIm(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi)$. Procedemos a ajustar un conjunto de ecuaciones de regresión a la media para los parámetros $\theta = [\omega, \eta, \beta, \phi, a_1, a_2, a_3]$

$$f(Y_i / \alpha_{0i}, \alpha_{1i}, \gamma_i, \phi) = \begin{cases} \alpha_{0i}(1-\gamma_i) & , si \ y=0 \\ c_i b\left(Y_i / \frac{\gamma_i(1-\alpha_{1i})}{1-\alpha_{0i}(1-\gamma_i)-\alpha_{1i}\gamma_i}, \phi\right) & , si \ y \in (0,1) \\ \alpha_{1i}\gamma_i & , si \ y=1 \end{cases}$$

$$\alpha_{0i} = g_{a_1}^{-1}(\tilde{x}_i^t \omega), \quad \alpha_{1i} = g_{a_2}^{-1}(\tilde{x}_i^t \eta), \quad \gamma_i = g_{a_3}^{-1}(x_i^t \beta), \quad c_i = 1 - \alpha_{0i}(1 - \gamma_i) - \alpha_{1i}\gamma_i,$$

$$\mu_i = \frac{\gamma_i(1 - \alpha_{1i})}{c_i}$$

Donde $\omega = (\omega_0, \omega_1, \omega_2, \dots, \omega_{k_1})$, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{k_2})$, $\eta = (\eta_0, \eta_1, \eta_2, \dots, \eta_{k_3})$ son vectores columna con los parámetros de la regresión $\tilde{x}_i = [1, x_i, \dots, x_{ik_1}]^t$, $\check{x}_i = [1, x_i, \dots, x_{ik_2}]^t$, $x_i = [1, x_i, \dots, x_{ik_3}]^t$ son vectores columna de las covariables por individuo i, y $g_{a_1}, g_{a_2}, g_{a_3}$ son

funciones de enlace y a_1, a_2, a_3 son los parámetros de potencia para cada una de las partes de la matriz de covariables.

Caso Log Log

$$\alpha_{0i} = g_{a_1}^{-1}(\tilde{x}_i^t \omega) = \exp(-a_1 \exp(\tilde{x}_i^t \omega))$$

$$\alpha_{1i} = g_{a_2}^{-1}(\tilde{x}_i^t \eta) = \exp(-a_2 \exp(\tilde{x}_i^t \eta))$$

$$\gamma_i = g_{a_3}^{-1}(x_i^t \beta) = \exp(-a_3 \exp(x_i^t \beta))$$

Caso LogLog complementario

$$\alpha_{0i} = g_{a_1}^{-1}(\tilde{x}_i^t \omega) = 1 - \exp(-a_1 \exp(\tilde{x}_i^t \omega))$$

$$\alpha_{1i} = g_{a_2}^{-1}(\tilde{x}_i^t \eta) = 1 - \exp(-a_2 \exp(\tilde{x}_i^t \eta))$$

$$\gamma_i = g_{a_3}^{-1}(x_i^t \beta) = 1 - \exp(-a_3 \exp(x_i^t \beta))$$

2.1.7.3 Función de verosimilitud del modelo

Por tanto, la función de verosimilitud, que depende del vector de parámetros desconocidos θ , viene dada por:

$$L(\theta) = f(Y_1 = y_1, \dots, Y_n = y_n / \theta) = \prod_{i=1}^n f(y_i / \theta)$$

Asumiendo que no se pierde generalidad, los valores dependientes para nuestra variable dependiente fraccionarias son ordenadas de 0,1 y (0,1) valores respectivamente, $\eta_0, n_1, n - m$ términos $m = \eta_0 + n_1$ nosotros podemos expresar la **función verosimilitud** como sigue:

$$L(\theta) = \prod_{i=1}^{n_0} \alpha_{0i} (1 - \gamma_i) \prod_{i=n_0+1}^m \alpha_{0i} (1 - \gamma_i) \prod_{i=n_0+1}^m c_i b(y_i / \mu_i, \phi)$$

2.1.7.4 Estimador de máxima verosimilitud del modelo

El estimador de máxima verosimilitud (EMV) para θ es el valor de $\hat{\theta}$ que maximiza la verosimilitud $L(\theta)$. Sin embargo, dado que el logaritmo es una función estrictamente positiva, la maximización se realiza sobre la función de log-verosimilitud definida por:

$$K(\theta) = \sum_{i=1}^{n_0} \log(\alpha_{0i}(1 - \gamma_i)) + \sum_{i=n_0+1}^m \log(\alpha_{1i}\gamma_i) + \sum_{i=m+1}^n \log(c_i) + \sum_{i=m+1}^n \log b(y_i/\mu_i, \phi)$$

Donde $K(\theta) = \log b(y_i/\mu_i, \phi)$ es igual:

$$\log \Gamma(\phi) - \log \Gamma(\mu_i\phi) - \log \Gamma((1 - \mu_i)\phi) + (\mu_i\phi - 1) \log(y_i) + ((1 - \mu_i)\phi - 1) \log(1 - y_i)$$

El término de la ecuación de máximo verosímil satisface: $g(\mu_i) = \frac{\partial \log b(y_i/\mu_i, \phi)}{\partial \mu_i} = \phi_i(Y_i^* - \mu_i^*)$

Donde $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$ $\mu_i^* = \psi(\mu_i\phi) - \psi((1 - \mu_i)\phi)$ siendo ψ la función digama i.e

donde $\psi(z) = \frac{\partial \log(\Gamma(z))}{\partial z}$, $z > 0$ Una propiedad muy conocida de la distribución beta y la función

de puntuación establece que el valor esperado de $\frac{\partial \log b(y_i/\mu_i, \phi)}{\partial \mu_i} = \phi_i(Y_i^* - \mu_i^*)$, es cero y por lo

tanto $E(Y_i^*) = \mu_i^*$

Los detalles de las primeras y segunda derivadas se dan en la parte final en el capítulo Anexos.

Debido al cálculo previo la matriz de información de Fisher estaría dada por $I(\theta) = -E\left(\frac{\partial^2 K(\theta)}{\partial \theta^2}\right)$

se obtiene directamente de la matriz Hessiana. La optimización de la función log-verosimilitud

puede llevarse a cabo mediante un algoritmo de puntuación de Fisher modificando o cualquier

algoritmo de optimización numérica. En este trabajo hemos utilizado una rutina de Matlab con el

vector de gradiente analítico y la matriz Hessiana dada al final del capítulo. Una vez obtenido el

estimado de máximo-verosimilitud θ se puede recuperar fácilmente a cada sujeto i , los

estimadores de máximo-verosimilitud de probabilidades en 0 y 1 la media de los valores continuos respectivamente. El vector gradiente y la matriz Hesiana se muestran a continuación:

$$\begin{aligned}
 \text{Gradiente} &= \left[\frac{\partial K(\theta)}{\partial \omega_j} \quad \frac{\partial K(\theta)}{\partial \eta_j} \quad \frac{\partial K(\theta)}{\partial \beta_j} \quad \frac{\partial K(\theta)}{\partial \phi} \quad \frac{\partial K(\theta)}{\partial a_1} \quad \frac{\partial K(\theta)}{\partial a_2} \quad \frac{\partial K(\theta)}{\partial a_3} \right] \\
 \text{Hesiana} &= \begin{bmatrix} \frac{\partial^2 K(\theta)}{\partial^2 \omega_j} & \frac{\partial K(\theta)}{\partial \omega_j \eta_h} & \frac{\partial K(\theta)}{\partial \omega_j \beta_h} & \frac{\partial K(\theta)}{\partial \omega_j \phi} & \frac{\partial K(\theta)}{\partial \omega_j a_1} & \frac{\partial K(\theta)}{\partial \omega_j a_2} & \frac{\partial K(\theta)}{\partial \omega_j a_3} \\ \cdot & \frac{\partial^2 K(\theta)}{\partial^2 \eta_j} & \frac{\partial K(\theta)}{\partial \eta_j \beta_h} & \frac{\partial K(\theta)}{\partial \eta_j \phi} & \frac{\partial K(\theta)}{\partial \eta_j a_1} & \frac{\partial K(\theta)}{\partial \eta_j a_2} & \frac{\partial K(\theta)}{\partial \eta_j a_3} \\ \cdot & \cdot & \frac{\partial^2 K(\theta)}{\partial^2 \beta_j} & \frac{\partial K(\theta)}{\partial \beta_j \phi} & \frac{\partial K(\theta)}{\partial \beta_j a_1} & \frac{\partial K(\theta)}{\partial \beta_j a_2} & \frac{\partial K(\theta)}{\partial \beta_j a_3} \\ \cdot & \cdot & \cdot & \frac{\partial^2 K(\theta)}{\partial^2 \phi} & \frac{\partial K(\theta)}{\partial \phi a_1} & \frac{\partial K(\theta)}{\partial \phi a_2} & \frac{\partial K(\theta)}{\partial \phi a_3} \\ \cdot & \cdot & \cdot & \cdot & \frac{\partial^2 K(\theta)}{\partial^2 a_1} & \frac{\partial K(\theta)}{\partial a_1 a_2} & \frac{\partial K(\theta)}{\partial a_1 a_3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\partial K^2(\theta)}{\partial^2 a_2} & \frac{\partial K(\theta)}{\partial a_2 a_3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \frac{\partial K^2(\theta)}{\partial^2 a_3} \end{bmatrix}
 \end{aligned}$$

2.1.7.5 Simulación de Datos

Se simuló 9 escenarios de datos, para un tamaño de muestra $n = 1000$ usando 500 simulaciones. Los escenarios son consecuencia de usar 3 parámetros de potencia y usar 3 casos de distribución de datos (0,1), 0 y 1 inflacionados correspondientes a los enlaces cloglog y loglog potencia. Los parámetros son los siguientes: $\theta = [\omega, \eta, \beta, \phi, a_1, a_2, a_3]$

Se ajustaron a cuatro covariables a lo largo de una función de enlace cloglog y loglog potencia. Las covariables se generaron a partir de una distribución multinormal con un vector de medias cero y matriz de varianza – covarianza y también asignamos valores iniciales al vector de parámetros, los cuales se muestran a continuación:

$$\Sigma = \begin{bmatrix} 0.4 & -0.5 & 0.4 & 0.2 \\ -0.5 & 1 & 0.2 & 0.1 \\ 0.4 & 0.2 & 3 & 1.5 \\ 0.2 & 0.1 & 1.5 & 1 \end{bmatrix}$$

$$Y_i \sim BIm(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi, a_1, a_2, a_3)$$

$$a_1 = [0.5, 1, 1.5]', \quad a_2 = [0.5, 1, 1.5]', \quad a_3 = [0.5, 1, 1.5]', \quad \phi = [10, 10, 10]'$$

$$\beta = [0.1, 0.2, -0.2, 0.4, 0.1]^T, \quad \eta = [0.1, -0.1, -0.2, 0.3, 0.5]^T,$$

$$\omega = [0.5, -0.1, -0.5, 0.5, 1]^T$$

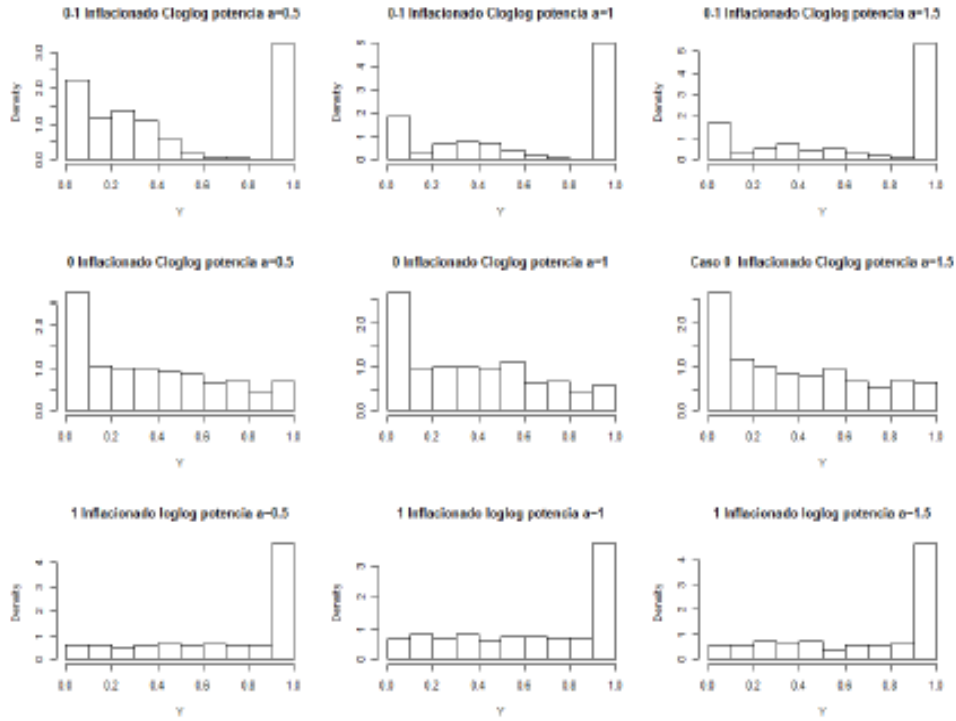
Mostramos el resumen de la simulación obtenida con los valores iniciales y tomando la distribución $Y_i \sim BIm(\alpha_{0i}, \alpha_{1i}, \gamma_i, \phi, a_1, a_2, a_3)$, con lo cual obtenemos 9 escenarios para el caso de parámetro potencia $a = (0.5, 1, 1.5)$ y para el caso 0 inflacionado, 1 inflacionado y 0-1 inflacionado, tener en cuenta que para el parámetro potencia 1, los 2 modelos tienen similares condiciones, los resultados se muestran a continuación (ver código de datos simulados en el capítulo de Anexos):

Tabla 1: Tabla de frecuencias para distribución de tres partes para la potencia $a=0.5$, $a=1.0$ y $a=1.5$

Caso	Potencia	% Frecuencia 0	% Frecuencia (0,1)	% Frecuencia 1
0-1 Inflacionado	0.50	18%	50%	32%
	1.00	16%	37%	47%
	1.50	13%	38%	49%
0 Inflacionado	0.50	21%	80%	0%
	1.00	19%	81%	0%
	1.50	18%	82%	0%
1-Inflacionado	0.50	0%	59%	42%
	1.00	0%	70%	30%
	1.50	0%	60%	40%

Fuente: Elaboración propia

Figura 3 : Distribuciones simuladas usando enlaces CLOGLOG y LOGLOG potencia $n=1000$



Fuente: Elaboración propia

2.1.4.6 Aplicación a los datos simulados: recuperación de parámetros

A continuación, mostramos los resultados para los 9 escenarios de datos simulados aplicando el modelo propuesto de enlaces asimétricos en un comparativo usando el enlace asimétrico de (Bayes & Valdivieso, 2016). A nivel genérico mostramos el indicador de AIC para mostrar el desempeño de ajuste para todos los escenarios en los modelos planteados; podemos ver que el mejor el modelo propuesto tiene un mejor ajuste en todos los escenarios simulados incluso en el escenario planteado solo por el modelo tradicional (potencia igual 1).

Tabla 2: Resumen de AIC para el modelo propuesto de enlaces asimétricos y logísticos para los 9 escenarios planteados en la simulación

Potencia	Caso	Modelo Tradicional	Modelo Propuesto
0.5	0 -1 Inflacionado	3.3775	2.5434
	0 Inflacionado	3.8951	2.4831
	1-Inflacionado	3.9936	2.1124
1	0 -1 Inflacionado	7.2867	4.4401
	0 Inflacionado	4.6508	3.3791
	1-Inflacionado	5.0937	2.9587
1.5	0 -1 Inflacionado	11.0719	9.5870
	0 Inflacionado	7.9010	6.8627
	1-Inflacionado	7.9716	6.7921

Fuente: Elaboración Propia

Nota: Los resultados del modelo tradicional es la aplicación de los valores en el modelo de (Bayes & Valdivieso, 2016)

Podemos generalizar que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste de parámetros, comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general todos los parámetros estimados esto para el caso del parámetro potencia $a = (0.5, 1, 1.5)$ y caso 0 inflacionado.

Tabla 3: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico $n=1000$ caso 0 inflacionado

Parámetro potencia "a"	Parámetros	Valor Real	Valor Estimado		Error Cuadrático Medio	
			mlogit	mcloglog	mlogit	mcloglog
0.5	α_1	0.50000	0.00000	0.47176	0.50000	0.10884
	α_3	0.50000	0.00000	0.24273	0.50000	0.30487
	ϕ	10.00000	2.29030	2.26580	7.71010	17.75000
	β_1	0.10000	-0.49252	0.01191	0.59319	0.26837
	β_2	0.20000	0.33468	0.16904	0.20866	0.16708
	β_3	-0.10000	0.10652	-0.07050	0.22558	0.10138
	β_4	0.40000	0.14484	0.18564	0.26018	0.19149
	β_5	0.10000	-0.10086	0.10050	0.21068	0.15388
	ω_1	0.10000	-0.82519	-0.08253	0.93152	0.57550
	ω_2	-0.10000	-0.43521	-0.06735	0.47342	0.19845
	ω_3	-0.20000	-0.23169	0.00758	0.24286	0.14770
	ω_4	0.30000	-0.04712	-0.04037	0.35406	0.34117
	ω_5	0.50000	0.38300	0.04710	0.17394	0.42097
	α_2	1.00000	0.00000	0.48142	1.00000	0.51860
	α_3	1.00000	0.00000	0.01037	1.00000	0.98963
1	ϕ	10.00000	2.05240	2.09030	7.94770	7.91000
	β_1	0.10000	-0.34126	-0.19994	0.44239	0.30532
	β_2	0.20000	0.11284	0.15207	0.13123	0.10099

	β_3	-0.10000	-0.02453	-0.08754	0.11748	0.08326
	β_4	0.40000	0.10716	0.19072	0.29374	0.20943
	β_5	0.10000	-0.06470	-0.00733	0.16726	0.11605
	ω_1	0.10000	-1.03680	0.53693	1.13880	0.44398
	ω_2	-0.10000	-0.34269	0.01770	0.28660	0.12431
	ω_3	-0.20000	-0.00890	0.09550	0.25719	0.30031
	ω_4	0.30000	-0.21953	-0.05844	0.52717	0.35875
	ω_5	0.50000	0.29714	-0.00348	0.21915	0.50423
	a_2	1.50000	0.00000	0.47851	1.50000	0.04782
	a_3	1.50000	0.00000	0.37124	1.50000	0.63382
	ϕ	10.00000	2.31320	2.31240	7.68740	7.68810
	β_1	0.10000	-0.49986	-0.00037	0.60105	0.12194
	β_2	0.20000	0.23994	0.15153	0.17097	0.14876
1.5	β_3	-0.10000	0.06188	0.02703	0.19056	0.15587
	β_4	0.40000	0.12118	0.17408	0.28599	0.23248
	β_5	0.10000	-0.10997	-0.03953	0.22747	0.16896
	ω_1	0.10000	-0.78374	-0.25420	0.88526	0.37829
	ω_2	-0.10000	-0.34484	-0.21049	0.37422	0.21677
	ω_3	-0.20000	-0.06997	-0.14465	0.16541	0.15101
	ω_4	0.30000	-0.03184	-0.01245	0.33677	0.31604
	ω_5	0.50000	0.35952	0.18107	0.18993	0.32378

Fuente: Elaboración propia, los resultados resaltados en plomo tienen un menor error cuadrático medio

Podemos generalizar que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste de parámetros, comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general todos los parámetros estimados esto para el caso del parámetro potencia $a = (0.5, 1, 1.5)$ y caso 0 inflacionado.

Tabla 4: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico $n=1000$ caso 1 inflacionado

Parámetro Potencia "a"	Parámetros	Valor Real	Valor Estimado		Error Cuadrático Medio	
			mlogit	mcloglog	mlogit	mcloglog
0.5	a_2	0.50000	0.00000	0.46725	0.50000	0.13575
	a_3	0.50000	0.00000	0.08784	0.50000	0.45790
	ϕ	10.00000	2.24560	2.16860	7.75460	7.83300
	β_1	0.10000	-0.47585	-0.02603	0.57650	0.17689
	β_2	0.20000	0.19162	0.10403	0.22742	0.15690
	β_3	-0.10000	-0.01749	-0.09349	0.12888	0.09578
	β_4	0.40000	0.15323	0.19950	0.25293	0.20855

	β_5	0.10000	-0.07118	0.07822	0.18369	0.16534
	η_1	0.10000	-0.73517	-0.01887	0.83683	0.16365
	η_2	-0.10000	-0.25948	-0.18097	0.37784	0.18154
	η_3	-0.20000	0.03627	-0.11689	0.27765	0.19811
	η_4	0.30000	-0.04699	-0.04282	0.36538	0.34385
	η_5	0.50000	-0.04745	0.26189	0.56523	0.25057
	a_2	1.00000	0.00000	0.49964	1.00000	0.30040
	a_3	1.00000	0.00000	0.04247	1.00000	0.75852
	ϕ	10.00000	2.13940	2.37510	7.86120	7.63030
	β_1	0.10000	-0.35003	-0.20023	0.45209	0.32128
	β_2	0.20000	0.19801	0.06773	0.10520	0.14411
	β_3	-0.10000	0.03046	-0.03861	0.15319	0.06425
1	β_4	0.40000	0.12595	0.12478	0.27596	0.27454
	β_5	0.10000	-0.13732	-0.03632	0.24592	0.13649
	η_1	0.10000	-0.86285	0.59114	0.96287	0.56935
	η_2	-0.10000	-0.58787	0.05333	0.61773	0.19135
	η_3	-0.20000	0.13754	-0.21051	0.34596	0.18968
	η_4	0.30000	-0.13789	-0.08338	0.43843	0.38454
	η_5	0.50000	-0.00891	0.33113	0.51513	0.21593
	a_2	1.50000	0.00000	0.50501	2.00000	1.49500
	a_3	1.50000	0.00000	0.35353	2.00000	1.64840
	ϕ	10.00000	2.35334	2.61261	8.64732	8.39333
	β_1	0.10000	-1.15590	-1.37510	1.40120	1.47750
	β_2	0.20000	-0.21777	-0.17758	0.51038	0.40488
	β_3	-0.10000	-0.14796	-0.05550	0.15264	0.10191
1.5	β_4	0.40000	-0.03965	0.03967	0.44212	0.37011
	β_5	0.10000	-0.05192	0.03778	0.20306	0.19037
	η_1	0.10000	0.44812	1.39850	1.42320	1.29870
	η_2	-0.10000	-0.14206	0.05397	0.19904	0.17508
	η_3	-0.20000	0.02294	0.01206	0.23221	0.21732
	η_4	0.30000	-0.09888	0.03444	0.43823	0.26600
	η_5	0.50000	-0.05149	-0.00350	0.56384	0.50530

Fuente: Elaboración propia, los resultados resaltados en plomo tienen un menor error cuadrático medio

Podemos generalizar que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste de parámetros, comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general todos los parámetros estimados esto para el caso del parámetro potencia $a = (0.5, 1, 1.5)$ y caso 0-1 inflacionado.

Tabla 5: Comparación de Error Cuadrático Medio para el modelo propuesto de enlaces asimétricos y logístico n=1000 caso 0-1 inflacionado

Parámetro Potencia "a"	Parámetros	Valor Real	Valor Estimado		Error Cuadrático Medio	
			mlogit	mcloglog	mlogit	mcloglog
0.5	α_1	0.50000	0.00000	0.50365	0.50000	0.01486
	α_2	0.50000	0.00000	0.48132	0.50000	0.10531
	α_3	0.50000	0.00000	0.46640	0.50000	0.20705
	ϕ	10.00000	2.35334	2.61261	8.64732	8.39333
	β_1	0.10000	-0.22711	-0.13417	0.33137	0.24262
	β_2	0.20000	0.31499	0.30280	0.23146	0.21850
	β_3	-0.10000	-0.12544	-0.11506	0.12140	0.10350
	β_4	0.40000	0.56965	0.32055	0.18622	0.10039
	β_5	0.10000	0.13693	-0.16534	0.11488	0.17697
	ω_1	0.10000	0.58501	0.52651	0.52310	0.47079
	ω_2	-0.10000	-0.29516	-0.26564	0.42252	0.38027
	ω_3	-0.20000	-0.30798	-0.27718	0.23246	0.20921
	ω_4	0.30000	0.53545	0.48191	0.28573	0.25716
	ω_5	0.50000	0.46945	0.42251	0.19519	0.17567
	η_1	0.50000	0.18069	0.17888	0.33983	0.33643
	η_2	-1.00000	-0.35934	-0.35575	0.74041	0.73301
	η_3	-0.50000	-0.36726	-0.36359	0.25072	0.24821
	η_4	0.50000	0.86367	0.85503	0.39564	0.39168
	η_5	1.00000	0.93175	0.92243	0.20862	0.20653
a_1	1.50000	0.00000	1.01520	1.50000	0.48750	
a_2	1.50000	0.00000	1.20610	1.50000	0.36988	
a_3	1.50000	0.00000	1.36170	1.50000	0.21245	
ϕ	10.00000	2.05240	2.09030	7.94770	7.91000	
1	b_1	0.10000	1.58680	-0.27347	1.48940	0.43315
	b_2	0.20000	0.27050	0.21913	0.27598	0.27418
	b_3	-0.10000	-0.22054	-0.06898	0.20095	0.17843
	b_4	0.40000	0.71938	0.42487	0.33222	0.14851
	b_5	0.10000	0.18908	-0.14364	0.15329	0.18780
	w_1	0.10000	0.16175	0.14558	0.74235	0.66812
	w_2	-0.10000	-0.84892	-0.76403	1.30750	1.17675
	w_3	-0.20000	-0.31413	-0.28272	0.67174	0.60457
w_4	0.30000	-0.05430	-0.04887	1.13210	1.01889	

	w5	0.50000	0.60836	0.54752	0.61652	0.55487
	n1	0.50000	0.63004	0.62374	0.28439	0.28155
	n2	-1.00000	-0.99957	-0.98957	0.36537	0.36172
	n3	-0.50000	-0.51526	-0.51011	0.24306	0.24063
	n4	0.50000	0.76199	0.75437	0.35932	0.35573
	n5	1.00000	1.34390	1.33046	0.44186	0.43744
	a1	1.00000	0.00000	1.04900	1.00000	0.07793
	a2	1.00000	0.00000	1.09450	1.00000	0.22972
	a3	1.00000	0.00000	1.22150	1.00000	1.99400
	phi	10.00000	2.13940	2.37510	7.86120	7.63030
	b1	0.10000	0.84458	-0.35975	0.74773	0.48740
	b2	0.20000	0.36348	0.29999	0.28160	0.12754
	b3	-0.10000	-0.15932	-0.00788	0.15363	0.12551
	b4	0.40000	0.65149	0.49176	0.26562	0.16860
	b5	0.10000	0.18200	-0.01389	0.15396	0.12265
1.5	w1	0.10000	-0.11194	-0.10075	1.67020	1.50318
	w2	-0.10000	-0.02883	-0.02594	3.37390	3.03651
	w3	-0.20000	-0.74606	-0.67145	3.36180	3.02562
	w4	0.30000	1.06150	0.95535	4.09860	3.68874
	w5	0.50000	0.84506	0.76055	1.38980	1.25082
	n1	0.50000	0.41511	0.41096	0.20372	0.20168
	n2	-1.00000	-0.74935	-0.74186	0.41523	0.41108
	n3	-0.50000	-0.48952	-0.48462	0.23095	0.22864
	n4	0.50000	0.79419	0.78625	0.35961	0.35601
	n5	1.00000	0.99377	0.98383	0.22313	0.22090

Fuente: Elaboración propia, los resultados resaltado en plomo tienen un menor error cuadrático medio

2.2 Marco Teórico relacionado al negocio

“A partir de la llegada de las transacciones a través de tarjeta de crédito en los años 60’s, el crédito de consumo se ha disparado desde la última mitad del siglo XX, lo que hace difícil el no incursionar en este producto ya que en muchos casos actualmente, solo se pueden hacer negociaciones a través de tarjetas de crédito como es el caso de compras a través del internet”. (Thomas, Endelman, & Crook, 2002)

“El producto tarjeta de crédito nace de la relación entre las instituciones financieras controladas y las franquicias que administran el mismo, cumpliendo con requerimientos mutuos de mantenimiento y control. En Perú se ha convertido en un instrumento altamente requerido por los clientes bancarios, debido al crecimiento sostenido de la demanda interna, de ahí la necesidad de incursionar en este segmento de mercado”. (Thomas, Endelman, & Crook, 2002)

“Los scores de cobranzas facilitan la implementación de mejores estrategias de recupero ya que consideran las características del producto, a través de modelos que permiten determinar la prioridad, el canal y el momento adecuados para cada cliente de manera de optimizar el uso de los recursos, minimizar los costos y maximizar el recupero”. (Solutions, 2014)

2.2.1 Riesgo de crédito

“El riesgo de crédito es el riesgo de que se incurra en una pérdida financiera si una contraparte de una transacción no cumple con sus obligaciones financieras de manera oportuna. Por lo tanto, es una función de lo siguiente: el valor de la posición expuesta al incumplimiento (la exposición al riesgo crediticio o crediticio); la proporción de este valor que se recuperaría en caso de incumplimiento; y la probabilidad de incumplimiento.

El riesgo de crédito también se utiliza de manera flexible para referirse a la probabilidad de incumplimiento, independientemente del valor que pueda perderse. Así también, otras definiciones de diferentes autores coinciden que el riesgo de crédito tiene relación con el deterioro en la capacidad de pago de los deudores originando una posible falta de atención total o

parcial del derecho de cobro llegado el vencimiento, lo que produciría una pérdida para el acreedor.” (Reyes Obispo & Leon Davan, 2014)

2.2.2 Riesgo país

“El riesgo país se refiere a una variedad de desafíos que se enfrentan al invertir en un país extranjero. Las corporaciones hacen un cálculo de riesgo-recompensa cuando deciden invertir en el extranjero. Para medir el riesgo país se puede dar mediante el riesgo económico, política financiera, riesgo político que está directamente ligado a al valor de los bonos soberanos”.

(Harvey, 2019)

2.2.3 Definiciones de estrategias de recupero en Banca

Para explicar la estrategia aplicada en la institución, partiremos de la definición la gestión de cobranzas:

Gestión de cobranzas

“El objetivo de la gestión de cobranzas es reactivar la relación comercial con el cliente, de manera que mantenga sus obligaciones al día. Inicia el momento que el deudor incumple, se puede realizar a través de procesos extrajudiciales: llamadas telefónicas, notificaciones, visitas y finalmente con un proceso judicial. Es parte del negocio bancario tener portafolios de clientes con atrasos o incobrables. El problema radica cuando este porcentaje sobrepasa los límites esperados, por lo que la estrategia de cobranza es parte integral del proceso de crédito, principalmente por dos razones: Mayor concentración en actividades de promoción y análisis de

créditos y el incremento en los niveles de morosidad institucionales”. (Reyes Obispo & Leon Davan, 2014)

2.2.4 Modelos de scoring de cobranzas

Herramienta estadística que permite predecir si un cliente incumplirá o no sus obligaciones de pago con una operación; y, cuantificar las probabilidades de recuperarse el saldo actual de la cartera en proceso de cobranza, predecir de forma estadística la solvencia del deudor y su deseo de pago, de manera que se observen resultados lógicos que permitan definir estrategias que reduzcan los saldos incobrables y los gastos operativos de recupero. Modelo predictivo cuya salida es la probabilidad de impago de la operación (probability of default, PD o π).

La estrategia de la gestión de cobranzas del Banco Peruano se aplica tanto para la banca corporativa como para la banca minorista; y dentro de este segmento el mercado masivo de la tarjeta de crédito. En el organigrama institucional del banco se mantiene un área responsable de la gestión de recupero y para la toma de decisiones cuenta con una comisión de cobranzas como órgano interno. La unidad de recuperaciones se apoya en empresas especializadas externas contratadas para la cobranza de la cartera de consumo principalmente la cartera masiva que incluye la tarjeta de crédito; y, adicionalmente coordina y controla los resultados de estas empresas. Pasa a este proceso, la cartera que cumple ciertas condiciones de la matriz de decisión que depende del factor “mora”. Existen varios tipos de cobranza: mora, pre mora y preventiva, extrajudicial; y, judicial. (Spongardi, 2002)

Cobranza Pre mora y Preventiva.

“Es la gestión enfocada a los productos de consumo en cuyo segmento se encuentra la tarjeta de crédito, objeto de esta investigación. Se remiten SMS y e-mails de acuerdo con el

comportamiento del cliente como campaña preventiva. Se realiza a través del departamento interno encargado de esta tarea, como una gerencia anticipada de la cartera por vencer hasta 7 días, notificando al cliente la próxima fecha de pago de su tarjeta y saldo.

Paralelamente se remite una carga a las mencionadas empresas de la información de la cartera de TC vencida (1 a 5 días), con ciertos parámetros de confidencialidad con el fin de cuidar el sigilo bancario. El área de recuperaciones recibe reportes diarios con los resultados de la gestión, de manera que se conozca el estado y se pueda proceder con los siguientes pasos en los casos requeridos”. (Solutions, 2014)

Cobranza extrajudicial

“Con las mismas empresas externas e incluidas la subsidiaria del grupo, se inicia la gestión de cobro extrajudicial a través de llamadas telefónicas, visitas domiciliarias, mensajes, cartas. Se aplica para la etapa de mora temprana y media, a partir de 1 hasta 180 días vencidos, en base al análisis de los ejecutivos de recuperaciones, mismo que se requiere reforzar con los resultados que arroje el modelo interno del score de cobranzas a desarrollar en esta investigación. El área de recuperaciones mantiene un panel de control de indicadores que permite medir la gestión de cada una de las empresas de cobranzas externas y evaluar comparativamente la oportunidad”. (Solutions, 2014)

Cobranza Judicial.

Con los resultados finales de la etapa anterior y una vez agotadas las gestiones extrajudiciales, el departamento de Recuperaciones y/o el Comité de Demanda de la subsidiaria resuelven iniciar acciones judiciales que aplica para la cartera con mora alta. Este departamento designa los clientes a entregar a cada abogado, distribuyendo de acuerdo con su infraestructura.

A la fecha de corte indicada en el párrafo anterior, la concentración de los informes es como sigue:

Canales de cobranza

- Cobranza de Campo (domiciliario)
- Cobranza de Call Center (Telefónico)
- Cobranza de Courier (Envío de cartas y mensajes de texto)

2.2.5 La gestión de cobranzas en banca

Según (Reyes Obispo & Leon Davan, 2014), la gestión de cobranza es el conjunto de acciones coordinadas, aplicadas adecuada y oportunamente a los clientes para lograr la recuperación de los créditos, de manera que los activos exigibles de la institución se conviertan en activos líquidos de la manera más rápida y eficiente posible, manteniendo en el proceso la buena disposición de los clientes para futuras negociaciones.

En ese sentido, la gestión de cobranza es un proceso bastante interactivo con los clientes, que parte del análisis de la situación del cliente, un oportuno y frecuente contacto con el cliente, ofreciendo en el proceso de negociación alternativas de solución oportunas para cada caso y registrando las acciones ejecutadas para realizar un seguimiento continuo y el control del cumplimiento de los acuerdos negociados. Algunas acciones típicas en la gestión de la cobranza se describen a continuación, junto con un flujo grama con el proceso que sigue la gestión de cobranza:

Análisis del caso: ¿Quién es el cliente? ¿Cuál es su situación? ¿Cuáles fueron las condiciones para el otorgamiento del crédito? ¿Por qué cayó en mora? Aquí podemos considerar fuentes internas y externas de información como centrales de riesgo, relación de deudores, etc.

Contacto con el cliente: ¿Qué información registra el cliente? ¿Dónde está ubicado el cliente? ¿Cuáles acciones ya fueron ejecutadas?

Diagnóstico: ¿Cuál es el problema a raíz de la morosidad actual? ¿Qué tipo de cliente tenemos?

Generación de alternativa: ¿Cuáles son las posibles soluciones? El objetivo de esta acción es la venta del beneficio para crear una cultura de pago en el cliente.

Obtención de compromisos de pago: ¿Realizamos una buena negociación? La IMF debe identificar claramente, cuándo, dónde, cómo y cuánto pagará el cliente, y recordar, por ejemplo, que el cliente que está en una situación de sobreendeudamiento o disminución de ingresos establecerá una jerarquía en el pago de las deudas.

Cumplimiento de compromisos de pago: ¿El cliente cumplió con el compromiso de pago en la fecha indicada? ¿Demuestra que quiere pagar? El objetivo es mostrar consistencia a lo largo de toda la gestión de cobranza. No basta el compromiso y la actitud positiva del cliente hacia el pago; los gestores de cobranza deben realizar seguimiento a los compromisos de pago.

Registro de Acciones: ¿Las acciones están siendo coordinadas? Es importante considerar o ponerse en el lugar de la persona que continuará la gestión de cobranza.

Seguimiento del caso: ¿Conocemos la actual situación del cliente y las acciones realizadas?

Intensificación de las acciones: ¿Cuál es la acción para tomar que nos permita recobrar el activo de manera más inmediata? ¿Cuáles son los activos que el cliente posee? ¿Qué podemos recuperar con una acción legal? En esta fase el interés es recuperar el activo aún a costa de perder al cliente.

2.2.6 Definición créditos perdida

“Es importante también que las IMF definan claramente las condiciones para reconocer la pérdida de un crédito, es decir - cuando la gestión de cobranza ha finalizado. Puede ser cuando se ha agotado todas las estrategias posibles para la recuperación de la deuda y/o cuando la probabilidad de pago es muy baja. En general, se debe analizar el costo-beneficio de las medidas judiciales, reportando el cliente moroso, y otras acciones permitidas por ley. El cliente debe percibir la gestión de cobranza como algo continuo y no esporádico, de allí que resulte muy importante el seguimiento oportuno y rápido de las acciones de cobranza por los diversos participantes de la gestión: call center, asesores de crédito, y gestores de campos. El cliente debe sentir que lo estamos marcando constantemente, siendo dinámicos, ágiles y resolutivos para controlar la situación; es fundamental realizar acciones dirigidas a todas las personas que participaron en el crédito, incluyendo cónyuges, garantes, y familiares/amigos que ofrecieron referencias, de acuerdo con su perfil de riesgo y probabilidad de pago.” (Castañeda Ulloa, 2015)

Capítulo III: Metodología de la investigación

En el presente capítulo se muestra los aspectos metodológicos considerados para llevar a cabo la tesis. Tales como la naturaleza de los datos, tipo de estudio, el preprocesamiento de los datos, técnicas estadísticas a emplear, entre otras.

3.1 Tipo, Nivel y Diseño de la investigación

3.1.1 Tipo de investigación

La presente investigación es de tipo teórica por que proponemos un nuevo modelo de regresión a la media para datos fraccionarios y también es de tipo aplicada cuantitativa ya que

quiere resolver un caso de estudio de predicción del ratio de recupero, que se genera en base al incumplimiento del pago de los clientes tarjetahabientes de un banco Peruano.

3.1.2 Nivel de investigación

El nivel de la investigación será descriptivo y explicativo.

3.1.3 Diseño de investigación

El diseño de la investigación es no experimental ya que observarán situaciones ya existentes, no se manipularán las variables solo se analizarán.

El diseño de Investigación no experimental será transeccional o transversal y correlacionales-causales ya que los datos serán analizados en un solo momento de tiempo y se analizara la relación entre las variables que como la afecta la variación entre ellas.

3.2 Población de estudio

La población de estudio corresponde al total de clientes de un banco peruano que han caído en mora usando el producto de tarjeta de crédito Z, esta población para el año 2017 asciende a 15,000 clientes.

3.3 Diseño muestral

La muestra fue seleccionada por conveniencia para los meses de Sep17, Oct17, Nov17 dado que la tarjeta de crédito Z salió al mercado en Julio 2017, los datos de mora se consideran estables 3 meses después de su lanzamiento. Se trabajará con una base de 12,519 clientes correspondientes en un periodo de corte de 3 meses (Sep17; Oct17; Nov17).

3.4 Técnicas e instrumento de recolección de datos

El banco peruano usa 2 instrumentos de recolección de datos:

- Un formulario donde ingresa los datos del cliente al momento de sacar la tarjeta, el cual es llenado por un asesor de ventas. Esta información se almacena en los sistemas de información del banco.
- Sistema de información interno ERP donde se almacena los datos de forma automática cuando el cliente empieza a usar su tarjeta de crédito, almacenando su consumo y movimientos de pago.

La información para trabajar la investigación será extraída de los sistemas de información del banco peruano, mediante queries de SQL usando los filtros adecuados en periodo, producto y estado de mora del cliente. Previamente se ha firmado un convenio entre las áreas de sistemas y cobranzas para acceder a los datos.

3.2 Operacionalización de variables

En tabla siguiente se muestra la matriz de consistencia, que relaciona los problemas, objetivos, hipótesis y las variables a considerar en el estudio de aplicación del modelo de enlaces asimétricos. La importancia de esta matriz se basa en la síntesis y consistencia de la investigación. Respecto a las variables en los datos proporcionado se encuentran una gran variedad, pero solo tomaremos las más representativas desde el lado de la completitud.

Tabla 6: Matriz de consistencia

Título: Regresión a la media para datos fraccionarios usando enlaces asimétricos con una aplicación a cobranzas de tarjetas de crédito				
Problemas	Objetivos	Hipótesis	Variables	
<p>General:</p> <p>¿Es posible desarrollar un modelo de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia y desarrollar un análisis comparativo con el modelo desarrollado por (Bayes & Valdivieso, 2016)?</p> <p>Específicos:</p> <p>¿Es posible realizar un análisis comparativo de desempeño usando datos de cobranzas a tarjetas de crédito, aplicando el modelo propuesto de enlaces asimétricos y el modelo desarrollado por (Bayes & Valdivieso, 2016)?</p> <p>¿Es posible identificar y cuantificar cuáles son las variables que tienen más influencia en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito?</p>	<p>General:</p> <p>Proponer un modelo de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia y realizar un análisis comparativo con el modelo desarrollado por (Bayes & Valdivieso, 2016)</p> <p>Específicos:</p> <p>Realizar un análisis comparativo usando datos de cobranzas a tarjetas de crédito, aplicando el modelo propuesto de enlaces asimétricos y el modelo desarrollado por (Bayes & Valdivieso, 2016).</p> <p>Identificar y cuantificar cuáles son las variables que tienen más influencia en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito.</p>	<p>General:</p> <p>El modelo propuesto de regresión a la media para una variable respuesta fraccionaria, usando enlaces asimétricos potencia tiene un menor ECM que el modelo desarrollado por (Bayes & Valdivieso, 2016) bajo el análisis de simulación.</p> <p>Específicos:</p> <p>El modelo propuesto de enlaces asimétricos tiene un menor ECM que el modelo desarrollado por (Bayes & Valdivieso, 2016) usando datos de cobranzas de una tarjeta de crédito.</p> <p>El peso de la importancia de las variables demográficas es mayor al de las variables de negocio en el modelo propuesto de enlaces asimétricos, aplicado a datos de cobranzas de tarjeta de crédito.</p>	<p>Dependiente:</p> <p>Y: Ratio de Recupero</p> <p>Independiente:</p> <p>X1: Variables de tipo financiera</p> <p>X2: Variables de tipo demográfica</p> <p>X3: Variables de tipo social</p> <p>X4: Rango de ingreso del cliente</p> <p>X5: Edad del cliente</p> <p>X6: Antigüedad en el banco</p> <p>X7: Número de entidades en el sistema financiero</p>	

Tabla 7 : Operacionalización de las variables

Rol	Nombre de la variable	Definición	Tipo de Variable	Valor
ID	ID_CORRELATIVO	ID Cliente correlativo	Cuantitativa	Valores de identificación
Predictora	FLG_BANCARIZADO	Flag Bancarizado (1) / no bancarizado (0)	Cualitativa	Toma valores de 0 y 1
Predictora	RANG_INGRESO	Ingreso estimado: rangos de menor hasta mayor ingreso (rang_ingreso_01, rang_ingreso_02, etc)	Cualitativa	Toma valores de rango_ingreso 01,02,03,04,05,06,07,08 y 09
Predictora	FLAG_LIMA_PROVINCIA	Lugar de residencia: Lima, Provincia	Cualitativa	Toma valores de Lima y Provincia
Predictora	EDAD	Edad	Cuantitativa	Toma valores positivos mayores a 18
Predictora	ANTIGÜEDAD	Antigüedad	Cuantitativa	Toma valores positivos mayores a 0
Predictora	RANG_SDO_PASIVO	Saldo ahorros: rangos de menor a mayor monto (cero, rango_sdo_01, rango_sdo_02, etc)	Cualitativa	Toma valores de cero, rango_sdo_01,02,03,04,05,06,07,08,09
Predictora	SDO_ACTIVO	Saldo / deuda en créditos	Cuantitativa	Toma todos los valores
Predictora	FLG_SEGURO	Flag tiene algún seguro: tiene seguro (1), no tiene seguro (0)	Cualitativa	Toma valores 0 y 1

Predictora	NRO_PRODUCTOS	Número de productos con la entidad: rangos de menor a mayor cantidad de productos (rango_01, rango_02, etc)	Cuantitativa	Toma valores mayores a 1
Predictora	NRO_ENTID_SSFF_MENOS0	Número de productos con la entidad: rangos de menor a mayor cantidad de productos (rango_01, rango_02, etc)	Cuantitativa	Toma valores mayores a 1 a 9
Predictora	FLG_NOMINA	Flag nómina: si (1), no (0)	Cualitativa	Toma valores de 0 y 1
Predictora	NRO_ACCES_CANAL1	Nro de operaciones por el canal 1	Cuantitativa	Toma valores positivos
Predictora	FLG_SDO_OTSSFF	Flag tiene saldo en otras entidades del sistema financiero: tiene saldo (1), no tiene saldo (0)	Cualitativa	Toma valores de 0 y 1
Respuesta	RECUPERO_MES1	Ratio de recupero del cliente luego de que ha caido en defaul en el mes 1	Cuantitativa	Toma valores positivos de tipo fracción <=0,1>
Respuesta	RECUPERO_MES2	Ratio de recupero del cliente luego de que ha caido en defaul en el mes 2	Cuantitativa	Toma valores positivos de tipo fracción <0,1>
Respuesta	RECUPERO_MES3	Ratio de recupero del cliente luego de que ha caido en defaul en el mes 3	Cuantitativa	Toma valores positivos de tipo fracción <0,1=>

Capítulo IV: Procesamiento de los datos

En esta fase analizamos los datos antes de ingresarlos al modelo planteado de enlaces asimétricos, procedemos con todos los pasos que considera la metodología de minería de datos previos al modelamiento Exploración, Limpieza y Transformación. Separamos la definición de variables en cuantitativas, cualitativas y ordinales mostrando los indicadores más relevantes para cada variable y un esbozo gráfico. Con lo cual nos ayudará a tener una idea de la posible distribución y dispersión de las variables que son candidatas para ingresar al modelo y sobre todo el análisis de la variable respuesta “Ratio de Recupero” para cada uno de los 3 meses.

4.1 Definición de variables cuantitativas

Tabla 8: Presentación descriptiva de variables cuantitativas

VARIABLES	N	Mínimo	Máximo	Media	Desv. típ.
EDAD	12519	18	91.0	32.7	9.0
ANTIGUEDAD	12519	0	19.0	3.9	3.3
SDO_ACTIVO_MENOS0	12519	0	11,348,500.0	20,422.4	165,830.5
SDO_ACTIVO_MENOS1	12519	0	11,289,560.0	20,013.0	165,201.7
SDO_ACTIVO_MENOS2	12519	0	11,302,990.0	19,554.4	165,176.6
SDO_ACTIVO_MENOS3	12519	0	10,706,290.0	18,995.6	162,116.1
SDO_ACTIVO_MENOS4	12519	0	10,620,280.0	18,017.1	139,595.6
SDO_ACTIVO_MENOS5	12519	0	10,791,890.0	17,648.3	149,297.6
NRO_ENTID_SSFF_MENOS0	12519	0.0	9.0	2.1	1.7
NRO_ENTID_SSFF_MENOS1	12519	0.0	9.0	2.1	1.7
NRO_ENTID_SSFF_MENOS2	12519	0.0	9.0	2.1	1.7
NRO_ENTID_SSFF_MENOS3	12519	0.0	9.0	2.0	1.7
NRO_ENTID_SSFF_MENOS4	12519	0.0	9.0	2.0	1.7
NRO_ENTID_SSFF_MENOS5	12519	0.0	9.0	2.0	1.7
RECUPERO_MES1	12519	0.00	1.0	0.4	0.3
RECUPERO_MES2	12519	0.00	1.0	0.5	0.3
RECUPERO_MES3	12519	.43	1.0	0.9	0.2

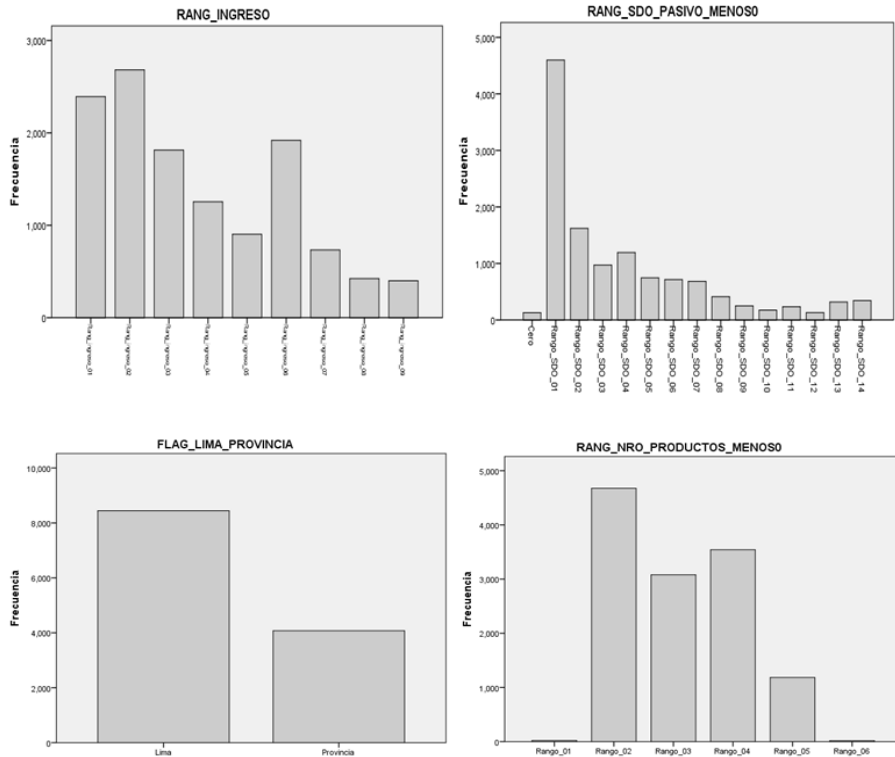
Fuente: Elaboración propia

El paso preliminar a todo estudio cuantitativo es la definición de las variables, en nuestro estudio la data análisis no presenta valores faltantes de ningún tipo como el caso de valores faltantes en las variables regresoras, valores faltantes en la variable de respuesta o valores faltantes en ambas. Por tal motivo no procederemos a realizar la imputación de valores.

4.2 Definición de variables cualitativas y ordinales

Se presenta las variables cualitativas y ordinales que son candidatas para ingresar al modelo en la sección de aplicación, tenemos: el rango de ingreso, rango de saldo, la zona y número de productos que se tienen con el banco. Para más detalles ver operacionalización de variables T12.

Figura 4 Presentación de variables cualitativas y ordinales



Fuente: Elaboración propia

Tabla 9: Presentación de variables cualitativas y ordinales

	Variable	Frecuencia	Porcentaje
Rango de Ingreso	Rang_ingreso_01	2,393	19.1
	Rang_ingreso_02	2,681	21.4
	Rang_ingreso_03	1,813	14.5
	Rang_ingreso_04	1,254	10
	Rang_ingreso_05	903	7.2
	Rang_ingreso_06	1,920	15.3
	Rang_ingreso_07	733	5.9
	Rang_ingreso_08	423	3.4
	Rang_ingreso_09	399	3.2
	Total	12,519	100
Zona residencia	Lima	8,445	67.5
	Provincia	4,074	32.5
	Total	12,519	100
Rango Saldo en Cuenta	Cero	129	1
	Rango_SDO_01	4,599	36.7
	Rango_SDO_02	1,621	12.9
	Rango_SDO_03	971	7.8
	Rango_SDO_04	1,193	9.5
	Rango_SDO_05	746	6
	Rango_SDO_06	715	5.7
	Rango_SDO_07	682	5.4
	Rango_SDO_08	412	3.3
	Rango_SDO_09	250	2
	Rango_SDO_10	174	1.4
	Rango_SDO_11	234	1.9
	Rango_SDO_12	130	1
	Rango_SDO_13	319	2.5
	Rango_SDO_14	344	2.7
Total	12,519	100	
Rango número de Productos	Rango_01	20	0.2
	Rango_02	4,676	37.4
	Rango_03	3,079	24.6
	Rango_04	3,542	28.3
	Rango_05	1,185	9.5
	Rango_06	17	0.1
	Total	12,519	100

4.3 Conversión de variables cualitativas (nominales y ordinales) a Dummy:

La conversión a variables Dummy busca estructurar una visión de la información clara sobre la Matriz de datos y que el algoritmo predictivo identifique estas variables del tipo categórico y no identifique estas variables categóricas como variables cuantitativas empleando el nivel de referencia de la categoría. Ver tabla con la conversión de variables en el Anexo C.

4.4 Revisión de valores atípicos

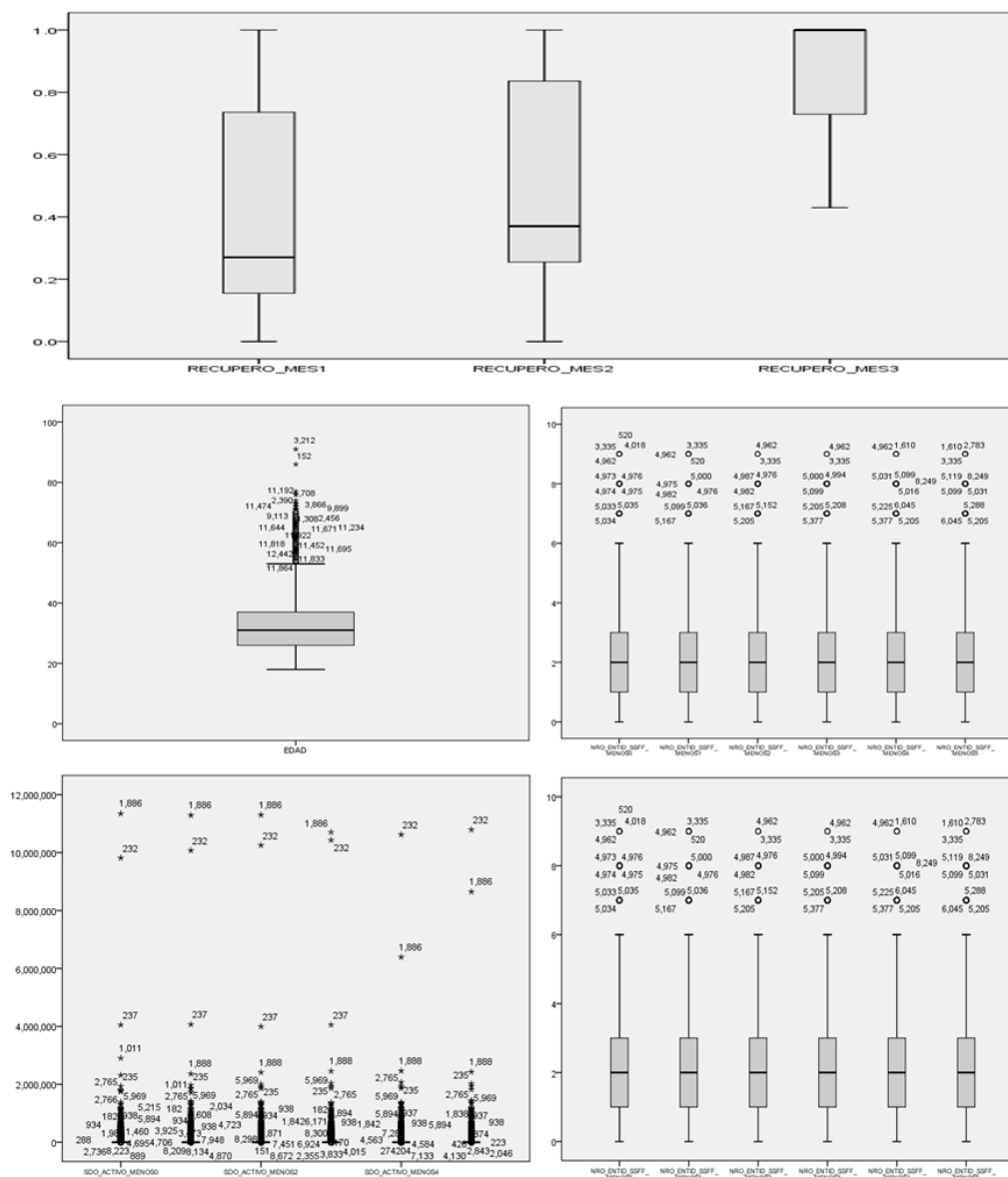
Ante un procedimiento de detección de valores atípicos debemos considerar no ser muy estrictos ni exhaustivos en la detección. Tengamos como premisa principal en el desarrollo de esta tarea, velar por mantener las relaciones entre variables que serán la estructura principal de un modelo predictivo. Dado que una eliminación de gran magnitud o sin revisar el impacto en los coeficientes de correlación, mantendremos un procedimiento consensado de técnicas de detección de valores atípicos, con la finalidad de identificar a los valores atípicos por comportamiento univariados y multivariados. Ya que no es necesario suponer que las variables, tengan una distribución normal. Aplicaremos la detección de casos ubicados fuera del rango intercuartílico, considerando que la información valiosa se encuentra dentro del rango.

Para nuestro caso hay que tener en cuenta que los enlaces asimétricos de tipo LOGLOG y CLOGLOG también son robustos ante valores atípicos, por lo cual se realizará el análisis, pero no eliminaremos los valores atípicos para probar la robustez del modelo.

4.5 Detección de atípicos univariados

Analizando de forma univariada a las variables vemos que todas presentan algún tipo de asimetría; la variable dependiente “ratio de recupero” es la única que no presenta valores atípicos, pero mientras pasan los meses la asimetría se marca aún más.

Figura 5 Detección de valores atípicos univariados



Fuente: Elaboración propia

4.6 Detección de atípicos multivariados

Por la sensibilidad a la detección de outlier que presenta el método de k-means, es muy utilizado en la detección de outlier multivariados. El procedimiento es solicitar un gran número de cluster (por ejemplo 50 cluster) y aquellos cluster con pocas observaciones y con mayor distancia entre los demás cluster serán los potenciales outliers multivariados. Los 50 clústeres fueron generados en el cuadro inferior, notaremos que elegimos que el punto de cortes para denominar al clúster como un grupo de outliers multivariado que tenga una frecuencia de menor o igual de 0.4% de los casos, tomando como limite el 1% que se agrupan en clusters atípicos, lo cuales marcamos con un color gris.

Tabla 10: Tabla de segmentos para identificación de outliers multivariados

Cluster	Frecuencia	Porcentaje
1	121	9.7
2	342	2.7
3	450	3.6
4	206	1.6
5	242	1.9
6	669	5.3
7	297	2.4
8	150	1.2
9	266	2.1
10	546	4.4
11	187	1.5
12	127	1.0
13	135	1.1
14	222	1.8
15	88	.7
16	84	.7
17	100	.8
18	127	1.0
19	59	.5
20	160	1.3

21	99	.8
22	84	.7
23	2	.0
24	38	.3
25	36	.3
26	524	4.2
27	242	1.9
28	103	.8
29	74	.6
30	340	2.7
31	252	2.0
32	494	3.9
33	262	2.1
34	457	3.7
35	281	2.2
36	258	2.1
37	172	1.4
38	215	1.7
39	79	.6
40	344	2.7
41	121	1.0
42	58	.5
43	168	1.3
44	157	1.3
45	184	1.5
46	47	.4
47	656	5.2
48	450	3.6
49	455	3.6
50	192	1.5
<hr/>		
Total	125	100.0
	19	

Fuente: Elaboración propia

4.7 Aplicación del modelo

Para este caso aplicamos el modelo de enlaces asimétrico desarrollado en el capítulo de marco teórico. Los resultados se muestran en el capitulo 4.

Capítulo V: Resultados

A continuación, mostraremos resultados de la aplicación del modelo planteado con enlaces asimétricos caso específico Cloglog y Loglog potencia. El objetivo es probar la hipótesis que el modelo planteado tiene un mejor ajuste que el modelo de (Bayes & Valdivieso, 2016) esto aplicado a los datos de cobranzas de un banco local peruano.

5.1 Resultados de ajuste de modelo

Para el caso del problema planteado, seleccionamos el enlace asimétrico LOGLOG potencia para el caso 0 inflacionado; para el mes 1 la distribución de los tarjeta habientes en estado de crédito default y consideramos como variables de entrada a Rango de Ingreso, edad, antigüedad en el banco y número de entidades financieras, para más detalles de la elección de estas variables ver (Deloitte, Abril 2017) y (Solutions, 2014), el porcentaje de pago se concentra a la izquierda con porcentajes altos de valores ceros. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico; obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico LOGLOG potencia tiene un mejor ajuste ante datos 0 inflacionados. Adicionalmente las variables ingresadas al modelo para el modelo propuesto son significativas en cualquiera de sus particiones, para el caso del enlace logístico hay algunas excepciones como rango de ingreso (η_2) y antigüedad en el banco (η_4).

Seleccionamos el enlace asimétrico LOGLOG potencia para el caso 0-1 inflacionado; para el mes 2 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas pesadas a los extremos 0 y 1 inflacionados. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico LOGLOG potencia tiene un mejor ajuste ante datos 0-1 inflacionados. Adicionalmente las variables ingresadas al modelo

para este caso son significativas en cualquiera de sus particiones excepto el rango de ingreso (ω_2, η_2) y antigüedad en el banco (ω_4, η_4).

Seleccionamos el enlace asimétrico CLOGLOG potencia para el caso 0-1 inflacionado; para el mes 2 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas pesadas a los extremos 0 y 1 inflacionados. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico CLOGLOG potencia tiene un mejor ajuste ante datos 0-1 inflacionados. Adicionalmente las variables ingresadas al modelo para el modelo propuesto son significativas en cualquiera de sus particiones, para el caso del enlace logístico hay algunas excepciones como antigüedad en el banco (η_4).

Finalmemnte, para el caso del problema planteado, seleccionamos el enlace asimétrico CLOGLOG potencia para el caso 1 inflacionado; para el mes 3 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas en valores 1 inflacionados. Realizamos un comparativo de resultados con el enlace clásico logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico CLOGLOG potencia tiene un mejor rendimiento ante distribuciones 1 inflacionados. Adicionalmente las variables ingresadas al modelo para este caso son significativas en cualquiera de sus particiones excepto antigüedad en el banco (ω_4).

Tabla 11: Comparación de resultados del modelo propuesto (loglog) y modelo enlaces logísticos para caso 0 inflacionada del mes1 y mes2

Parámetros	Covariable	Mes1 loglog		Mes2 loglog	
		mloglog (SE)	mlogistica(SE)	mloglog (SE)	mlogistica(SE)
a_1				0.89973(0.000000)**	
a_2		0.31511(0.00001)**		0.90041(0.000000)**	
a_3		1.5469(0.00003)**		0.90887(0.000000)**	
ϕ		20.395(0.02104)**	2.8734(0.00013)**	8.6778(0.000011)**	3.7341(0.002264)**
β_1	Intercepto	0.010475(0.000027)**	-0.88877(0.000008)**	0.26802(0.000002)**	-0.3239(0.000014)**
β_2	RANGO_INGRESO	6.8355(0.00000)**	9.6891(0.000662)**	7.4245(0.000012)**	12.541(0.002425)**
β_3	EDAD	-0.38749(0.000003)**	-0.99307(0.000013)**	-0.87076(0.000009)**	-0.9283(0.000012)**
β_4	ANTIGUEDAD_BANCO	5.9014(0.000000)**	8.4751(0.000541)**	4.7826(0.000194)**	11.272(0.001939)**
β_5	NRO_ENTID_SFFF	7.443(0.000000)**	7.239(0.000002)**	2.4225(0.000166)**	5.9032(0.000331)**
ω_1	Intercepto			-3.3119(0.000000)**	10.977(0.016422)**
ω_2	RANGO_INGRESO			-64.313(0.000000)**	-1000(69.927000)
ω_3	EDAD			2.9943(0.000000)**	0.35783(0.000527)**
ω_4	ANTIGUEDAD_BANCO			-64.271(0.045483)**	-1000(73.553000)
ω_5	NRO_ENTID_SFFF			-2.9886(0.000000)**	-8.7614(0.002623)**
η_1	Intercepto	-2.4692(0.001698)**	8.5442(0.003274)**	-1.6732(0.000009)**	-3.6854(0.000445)**
η_2	RANGO_INGRESO	4.716(0.001054)**	-475.71(18.71700)	-89.897(0.000009)**	14.263(0.864840)
η_3	EDAD	-5.0803(0.001613)**	0.77301(0.000147)**	1.8107(0.000055)**	-0.28316(0.000127)**
η_4	ANTIGUEDAD_BANCO	4.9352(0.000777)**	-1000(81.171000)	-35.424(0.000043)**	18.452(0.238270)
η_5	NRO_ENTID_SFFF	-3.2436(0.0009597)**	-9.4476(0.007469)	3.0458(0.000039)**	8.5337(0.003063)**

Fuente: Elaboración propia

Tabla 12: Comparación del modelo propuesto (cloglog) y enlaces logísticos para caso 0 inflacionada del mes2 y mes3

Parámetros	Covariable	Mes2 cloglog		Mes3 cloglog	
		mcloglog (SE)	mlogistica(SE)	mcloglog (SE)	mlogistica(SE)
α_1		1.0486(0.000024)**			
α_2		0.25967(0.000004)**		0.90113(0.000000)**	
α_3		0.18604(0.000007)**		0.90162(0.000000)**	
ϕ		9.0475(0.0000000)**	3.7341(0.002264)**	12.459(0.000028)**	3.7187(0.005290)**
β_1	Intercepto	-0.23518(0.000009)**	-0.3239(0.000014)**	0.58997(0.000000)**	0.38755(0.000002)**
β_2	RANGO_INGRESO	7.3206(0.000006)**	12.541(0.002425)**	-1.412(0.001875)**	5.3617(0.000296)**
β_3	EDAD	-0.85529(0.000008)**	-0.9283(0.000012)**	0.29528(0.000008)**	0.0033683(0.000000)**
β_4	ANTIGUEDAD_BANCO	6.6964(0.000009)**	11.272(0.001939)**	-3.7536(0.003396)**	4.8632(0.000350)**
β_5	NRO_ENTID_SSFF	4.1579(0.000006)**	5.9032(0.000331)**	-0.34304(0.000596)**	3.7173(0.000141)**
ω_1	Intercepto	-2.4329(0.000068)**	10.977(0.016422)**	-1.7151(0.000216)**	-3.7937(0.000014)**
ω_2	RANGO_INGRESO	-63.383(0.000076)**	-1000(69.9270000)	-75.922(0.010674)**	12.943(0.477360)
ω_3	EDAD	3.8511(0.000067)**	0.35783(0.000527)**	5.2009(0.000412)**	0.045119(0.000664)**
ω_4	ANTIGUEDAD_BANCO	-40.38(0.000000)**	-1000(73.553000)	-53.195(0.013058)**	16.637(0.259940)
ω_5	NRO_ENTID_SSFF	-2.0993(0.000069)**	-8.7614(0.002623)**	-6.8292(0.001152)**	9.2091(0.011966)**
η_1	Intercepto	-2.7674(0.000011)**	-3.6854(0.000134)**		
η_2	RANGO_INGRESO	16.468(0.000011)**	14.263(0.000534)**		
η_3	EDAD	-0.40595(0.000011)**	-0.28316(0.000005)**		
η_4	ANTIGUEDAD_BANCO	17.434(0.000011)**	18.452(0.000032)**		
η_5	NRO_ENTID_SSFF	9.9988(0.000011)**	8.5337(0.000271)**		

Fuente: Elaboración propia

Coefficientes estimados para los modelos de regresión a la media usando enlaces asimétricos 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 **).

Tabla 13: Comparación de AIC, BIC para metodología de enlaces asimétricos y enlaces logísticos caso 0,1 y 0-1 inflacionado Mes1 a 3

	Modelo	AIC	BIC
Mes 1	mloglog	3.0371	92.2571
	mlogistico	10.3191	110.9742
Mes 2	mloglog	15.8585	142.2535
	mlogistico	16.3410	163.6060
Mes 2	mcloglog	14.0734	140.4685
	mlogistico	16.3410	163.6060
Mes 3	mcloglog	4.6883	93.9083
	mlogistico	15.1126	115.7676

Fuente: Elaboración propia

5.1.1 Resultados predictivos del modelo

Se realizó la selección de la muestra de validación y de entrenamiento, tomando un 20% para muestra de validación y un 80% para entrenamiento del modelo, con lo cual garantizamos una representatividad para la evaluación del poder predictivo.

Tabla 14: Tamaños de muestra para entrenamiento y validación del modelo

Base Total	12,519
Muestra entrenamiento	10,016
Muestra validación	2503

Fuente: Elaboración propia

Se aplicó el modelo planteado con enlaces asimétricos caso específico usando enlaces Cloglog y Loglog potencia. El objetivo es probar la hipótesis que el modelo planteado tiene un

mejor rendimiento predictivo que el modelo de (Bayes & Valdivieso, 2016) esto aplicado a los datos de cobranzas de un banco local peruano. A continuación, se muestra la tabla 14, donde se muestran los indicadores que miden el rendimiento del modelo tales como ECM (error cuadrático medio), R-Cuadrado (coeficiente de determinación) y MAPE (Error porcentual absoluto medio), la parte teórica del cálculo de los indicadores la podemos encontrar en la sección de Marco Teórico.

Tabla 15: Indicadores ECM y MAPE para la muestra de validación

	Modelo (enlace)	ECM	MAPE
Mes 1	Mloglogp (0 inflacionada)	0.1136	0.1210
	Mlogistico(0 inflacionada)	0.1740	0.1991
Mes 2	Mloglogp (0-1 inflacionada)	0.1954	0.1722
	Mlogistico (0-1 inflacionada)	0.2623	0.1879
Mes 2	Mcloglog(0-1 inflacionada)	0.2499	0.1737
	Mlogistico(0-1 inflacionada)	0.2623	0.1877
Mes 3	Mcloglog (1 inflacionada)	0.1597	0.1912
	mlogistico(1 inflacionada)	0.2248	0.1939

Fuente: Elaboración propia

Luego del cálculo de indicadores que miden el desempeño de predicción del modelo, observamos a nivel general que el modelo propuesto de enlaces asimétricos es mejor que el modelo clásico planteado por (Bayes & Valdivieso, 2016). Esto ocurre para los 3 casos donde la distribución de la variable recupero es representada 0, 1 0-1 inflacionada. Otro punto a resaltar es que el modelo tiene una mejor predicción para el ratio de recupero del mes 1 (0 inflacionadaos), esto se da al analizar los 2 indicadores de evaluación del modelo ECM (Deloitte, Abril 2017) y MAPE.

Capítulo VI: Conclusiones

6.1 Aplicación del modelo a datos simulados

Se generó una simulación con 9 escenarios para datos para tamaños de muestra $N=1000$ usando 500 simulaciones. Los escenarios son consecuencia de usar 3 parámetros de potencia y usar casos de distribución de datos (0,1) ,0 y 1 inflacionados respectivamente y se realiza un comparativo con el modelo usando enlaces logísticos de (Bayes & Valdivieso, 2016). Los detalles del código para generación de la muestra se de en capítulo de Anexo 9.3.

Podemos generalizar para el caso 0 inflacionado, que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste en la estimación de parámetros (β, ω, η) , comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general para todos los parámetros estimados, esto para el caso del parámetro potencia $a= (0.5, 1, 1.5)$.

Podemos generalizar para el caso 0-1 inflacionado, que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste de parámetros (β, ω, η) , comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general todos los parámetros estimados esto para el caso del parámetro potencia $a= (0.5, 1, 1.5)$.

Podemos generalizar para el caso 1 inflacionado, que el modelo propuesto usando enlaces asimétricos tienen un mejor ajuste de parámetros (β, ω, η) , comparándolo con el modelo aplicando de enlace logístico de (Bayes & Valdivieso, 2016). Se obtiene un menor error cuadrático medio en general todos los parámetros estimados esto para el caso del parámetro potencia $a= (0.5, 1, 1.5)$.

Luego del análisis realizado en los capítulos anterior, tenemos suficiente evidencia estadística para afirmar que el modelo propuesto usando enlaces asimétricos en particular los enlaces cloglog y loglog potencia son más eficientes que el modelo tradicional usando el enlace Logístico (Bayes & Valdivieso, 2016); tomando como aplicación los datos simulados. Indicadores analizados para evaluar el ajuste del modelo es ECM (error cuadrático medio).

6.2 Aplicación del modelo a datos reales

Para el caso del problema planteado, seleccionamos el enlace asimétrico loglog potencia para el caso 0 inflacionado; para el mes 1 la distribución de las tarjetas habientes en estado de crédito default; el porcentaje de pago se concentra a la izquierda con porcentajes altos de valores 0 inflacionados. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico; obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico loglog potencia tiene un mejor ajuste ante datos 0 inflacionados. Adicionalmente las variables ingresadas al modelo propuesto son significativas en cualquiera de sus particiones, para el caso del enlace logístico hay algunas excepciones como rango de ingreso (η_2) y antigüedad en el banco (η_4).

Seleccionamos el enlace asimétrico loglog potencia para el caso 0-1 inflacionado; para el mes 2 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas pesadas a los extremos 0 y 1 inflacionados. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico loglog potencia tiene un mejor ajuste ante datos 0-1 inflacionados. Adicionalmente las variables ingresadas al modelo para este caso son significativas en cualquiera de sus particiones excepto el rango de ingreso (ω_2, η_2) y antigüedad en el banco (ω_4, η_4).

Seleccionamos el enlace asimétrico cloglog potencia para el caso 0-1 inflacionado; para el mes 2 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas pesadas a los extremos 0 y 1 inflacionados. Realizamos un comparativo con el modelo planteado por (Bayes & Valdivieso, 2016) usando el enlace logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico cloglog potencia tiene un mejor ajuste ante datos 0-1 inflacionados. Adicionalmente las variables ingresadas al modelo para el modelo propuesto son significativas en cualquiera de sus particiones, para el caso del enlace logístico hay algunas excepciones como antigüedad en el banco (η_4).

Finalmemnte para el caso del problema planteado, seleccionamos el enlace asimétrico cloglog potencia para el caso 1 inflacionado; para el mes 3 la distribución de las tarjetas habientes en estado de crédito default, el porcentaje de pago se concentra con colas en valores 1 inflacionados. Realizamos un comparativo de resultados con el enlace clásico logístico obteniendo un AIC y BIC mayores, lo cual muestra el enlace asimétrico cloglog potencia tiene un mejor rendimiento ante distribuciones 1 inflacionados. Adicionalmente las variables ingresadas al modelo para este caso son significativas en cualquiera de sus particiones excepto antigüedad en el banco (ω_4).

Luego del cálculo de indicadores que miden el desempeño de predicción del modelo esto calculado sobre datos de validación, observamos a nivel general que el modelo propuesto de enlaces asimétricos es mejor que el modelo clásico planteado por (Bayes & Valdivieso, 2016). Esto ocurre para los 3 casos donde la distribución de la variable recupero es representada 0, 1 0-1 inflacionada. Otro punto a resaltar es que el modelo tiene una mejor predicción para el ratio de recupero del mes 1 (0 inflacionadaos), esto se da al analizar los 2 indicadores de evaluación del modelo ECM (Deloitte, Abril 2017) y MAPE.

Es posible proponer un modelo de regresión a la media para una variable respuesta fraccionaria usando enlaces asimétricos de tipo loglog y cloglog potencia, los cuales son generados por distribuciones Gumbel Exponenciadas.

Luego del análisis desarrollado en los capítulos anteriores, existe suficiente evidencia estadística para afirmar que el modelo propuesto usando enlaces asimétricos en particular los enlaces cloglog y loglog es más eficiente que el modelo usando el enlace Logístico (Bayes & Valdivieso, 2016) tomando como aplicación los datos de un banco peruano específicamente aplicado a un área de cobranzas. Indicadores usados para medir el ajuste del modelo son AIC y BIC; y los indicadores usados para medir el rendimiento predictivo del modelo son ECM y MAPE.

Capítulo VII: Recomendaciones

De forma general se recomienda realizar extensiones tomando como base la investigación de (Bayes & Valdivieso, 2016) y (Papke & Wooldridge, 1996) los cuales se tomaron como referencia principal para desarrollar esta investigación:

- Ampliar los enlaces aplicados a la media, para la presente investigación se tomó el enlace LOGLOG y CLOGLOG potencia, el cual se podría extender a usar enlaces que desprenden de las distribuciones GEV (distribución de valores extremos generalizados)
- Replantear el modelo propuesto desde la mirada de la estadística Bayesiana, dado que lo planteado se realizó usando estadística clásica, en particular se halló los parámetros estimados tomando la maximización de la función de log-verosimilitud, usando una rutina de cálculo de gradiente analítico y matriz de información de Fisher.
- Realizar extensiones de la metodología planteada por (Papke & Wooldridge, 1996), tomando una distribución diferente a la Beta, pero que soporte la lectura de valores positivos entre 0,1, tales como la distribución SIMPLEX, posterior a eso podría desprenderse extensiones al tomar diferentes enlaces como son los logit, probit, loglog y cloglog.

Referencias Bibliográficas

- Ramalho, J., & da Silva, J. (2009). A two-part fractional regression model for the. *Quantitative Finance*, 621-636.
- A beta inflated mean regression model with mixed effects for fractional response variables. (2017). *Tesis PUCP*.
- Albeverio, S., Jentsch, V., & Kantz, H. (2005). *Extreme Events in Nature and Society*. Springer. ISBN: 978-3540286103.
- Bayes, C., & Valdivieso, L. (2016). *A beta inflated mean regression model for fractional response*. Lima: Journal of applied statistics.
- Bayes, C., Bazán, J., & García, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*, 841-866.
- Calabrese, R. (2012). *Regression Model for Proportions with Probability Masses at Zero and One*. Dublin: Geary Dynamic Lab, Geary Institute University College Dublin.
- Castañeda Ulloa, H. (2015). *Factores de riesgo en morosidad de tarjetas de credito del banco BBVA continental*. Trujillo Perú.
- Christopher J, S., Melguizo Castro, M., & Zoran , B. (2011). Modeling Percentage Outcomes: The Beta_Regression Macro. *SAS Global Forum 2011*, 335-2011.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer. ISBN:978-1852334598.
- Cribari-Neto, F., & Ferrari, S. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31(7), 799-815.
- Deloitte, C. (Abril 2017). Model Risk Management. *Driving the value in modelling*. Madrid.

- Flores, J. C. (2018). modelos de regresion a la media simplex inflacionada para proporciones. *PUCP Tesis*.
- Gupta, R. D., & Kundu, D. (1999). Generalized Exponential Distributions. *Austral & New Zealand J. Statistics*, 173-188.
- Haan, L., & Ferreira, A. (2006). *Extreme Value Theory. An Introduction*. Springer. ISBN: 978-0387239460.
- Kotz, S., & Nadarajah, S. (2001). *Extreme Value Distributions. Theory and Applications*. World Scientific. ISBN: 978-1860942242.
- Leadbetter, M., Lindgren, G., & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. First Edition. Springer. ISBN: 978-0387907314.
- MONTALVAN, E. C. (2014). *Prediccion de ambientes de posicion sedimentaria en base a caracteristica batimetricas, morfologicas sedimentarias en el margen continental peruano mediante modelos de estadistica espacial*. Lima: TESIS UNI.
- Nadarajah, S. (2005). The exponentiated Gumbel distribution with climate application. *Wiley InterScience (www.interscience.wiley.com)*, *Environmetrics* 2006; 17: 13–23.
- Ospina, R., & Ferrari, S. (2010). Inflated beta distributions. *Statistical Papers*, 111-126.
- Ospina, R., & Ferrari, S. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, 1609-1623.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric Methods for fractional response variables with an application to 401 (K) plan participation rate. *Journal of applied econometrics*, VOL. 11,619-632.
- Peussa, A. (2016). *Credit risk scorecard estimation by logistic*. Department of mathematics and statistics university of helsinki .

- Reyes Obispo, A., & Leon Davan, D. (2014). *Capacidad predictiva de los modelos de maquina de vectores de soporte y modelos de regresión logistica en el analisis de riesgo de credito*. Lima: TESIS UNI.
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods* 11, 54.
- Smithson, M., & Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables. *American Psychological Association*, Vol. 11, No. 1, 54–71.
- Solutions, M. (2014). Driving the value in modelling. *Driving the value in modelling*. Mexico DF.
- Spongardi, M. (2002). *Calidad crediticia de los usuarios de una tarjeta de credito*. Mar del plata: Facultad de ciencias economia y soales.
- Thomas, L., Endelman, D., & Crook, J. (2002). *Credit Scoring and Its Applications*. Society for industrial and applied mathematics .
- Wang, X., & Dipak, K. (2010). Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption. *The Annals of Applied Statistics*, Vol. 4, No. 4, 2000–2023.

Anexos

9.1 Cálculo Derivadas parciales y dobles

Mostramos las derivadas parciales y las segundas derivadas para hallar la matriz Hesiana y el gradiente analítico.

Tomando las derivadas parciales

$$\frac{\partial \alpha_{0i}}{\partial \omega_j} = \frac{1}{g_{a_1}^t(g_{a_1}^{-1}(\tilde{x}_i^t \omega))} = \frac{\tilde{x}_{ij}}{g_{a_1}^t(\alpha_{0i})}$$

$$\frac{\partial \alpha_{1i}}{\partial \eta_j} = \frac{1}{g_{a_2}^t(g_{a_2}^{-1}(\tilde{x}_i^t \eta))} \tilde{x}_{ij} = \frac{\tilde{x}_{ij}}{g_{a_2}^t(\alpha_{1i})}$$

$$\frac{\partial \gamma_{1i}}{\partial \beta_j} = \frac{1}{g_{a_3}^t(g_{a_3}^{-1}(x_i^t \beta))} x_{ij} = \frac{x_{ij}}{g_{a_3}^t(\gamma_i)}$$

$$\frac{\partial \alpha_{0i}}{\partial a_1} = (1 - \alpha_{0i}) \exp(\tilde{x}_i^t \omega)$$

$$\frac{\partial \alpha_{1i}}{\partial a_2} = (1 - \alpha_{1i}) \exp(\tilde{x}_i^t \eta)$$

$$\frac{\partial \gamma_i}{\partial a_3} = (1 - \gamma_i) \exp(x_i^t \beta)$$

$$\frac{\partial \mu_i}{\partial \omega_j} = \frac{\partial \mu_i}{\partial \alpha_{0i}} \frac{\partial \alpha_{0i}}{\partial \omega_j} = \frac{\gamma_i(1 - \gamma_i)(1 - \alpha_{1i})}{c_i^2 g_{a_1}^t(\alpha_{0i})} \tilde{x}_{ij}$$

$$\frac{\partial \mu_i}{\partial \eta_j} = \frac{\partial \mu_i}{\partial \alpha_{1i}} \frac{\partial \alpha_{1i}}{\partial \eta_j} = -\frac{\gamma_i(1 - \gamma_i)(1 - \alpha_{0i})}{c_i^2 g_{a_2}^t(\alpha_{1i})} \tilde{x}_{ij}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial \beta_j} = \frac{(1 - \alpha_{0i})(1 - \alpha_{1i})}{c_i^2 g_{a_3}^t(\gamma_i)} x_{ij}$$

$$\frac{\partial \mu_i}{\partial \alpha_{0i}} = \frac{\partial \mu_i}{\partial \alpha_{0i}} \frac{\partial \alpha_{0i}}{\partial a_1} = \frac{\gamma_i(1 - \alpha_{0i})(1 - \alpha_{1i})}{c_i^2} \exp(\tilde{x}_i^t \omega)$$

$$\begin{aligned} \frac{\partial \mu_i}{\partial a_1} &= \frac{\partial \mu_i}{\partial \alpha_{0i}} \frac{\partial \alpha_{0i}}{\partial a_1} = \frac{\gamma_i(1-\gamma_i)(1-\alpha_{1i})}{c_i^2} \alpha_{0i} \exp(-g_1^t(\alpha_{0i})) \\ * \frac{\partial \mu_i}{\partial a_1} &= \frac{\partial \mu_i}{\partial \alpha_{0i}} \frac{\partial \alpha_{0i}}{\partial a_1} = \frac{\gamma_i(1-\gamma_i)(1-\alpha_{1i})}{c_i^2} (1-\alpha_{0i}) \exp(g_1^t(\alpha_{0i})) \\ \frac{\partial \mu_i}{\partial a_2} &= \frac{\partial \mu_i}{\partial \alpha_{1i}} \frac{\partial \alpha_{1i}}{\partial a_2} = -\frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i})}{c_i^2} \alpha_{1i} \exp(-g_2^t(\alpha_{1i})) \\ * \frac{\partial \mu_i}{\partial a_2} &= \frac{\partial \mu_i}{\partial \alpha_{1i}} \frac{\partial \alpha_{1i}}{\partial a_2} = -\frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i})}{c_i^2} (1-\alpha_{1i}) \exp(g_2^t(\alpha_{1i})) \\ \frac{\partial \mu_i}{\partial a_3} &= \frac{\partial \mu_i}{\partial \gamma_{1i}} \frac{\partial \gamma_{1i}}{\partial a_3} = -\frac{(1-\alpha_{1i})(1-\alpha_{0i})}{c_i^2} \gamma_i \exp(-g_3^t(\gamma_i)) \\ * \frac{\partial \mu_i}{\partial a_3} &= \frac{\partial \mu_i}{\partial \gamma_{1i}} \frac{\partial \gamma_{1i}}{\partial a_3} = \frac{(1-\alpha_{1i})(1-\alpha_{0i})}{c_i^2} (1-\gamma_i) \exp(\gamma_i) \end{aligned}$$

$$\frac{\partial \mu_i}{\partial \alpha_{0i}} = \frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i})}{c_i^2}$$

$$\frac{\partial \mu_i}{\partial \alpha_{1i}} = -\frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i})}{c_i^2}$$

$$\frac{\partial \mu_i}{\partial \gamma_i} = \frac{(1-\alpha_{0i})(1-\alpha_{1i})}{c_i^2}$$

$$\frac{\partial \mu_i^*}{\partial \mu_i} = \phi H_i$$

$$\frac{\partial c_i}{\partial \alpha_{0i}} = \gamma_i - 1$$

$$\frac{\partial c_i}{\partial \alpha_{1i}} = -\gamma_i$$

$$\frac{\partial c_i}{\partial \gamma_i} = \alpha_{0i} - \alpha_{1i}$$

$$\frac{\partial c_i}{\partial \omega_i} =$$

$$(\gamma_i - 1) \frac{x_{ij}}{g_1^t(\alpha_{0i})}$$

$$\frac{\partial c_i}{\partial n_j} = -\gamma_i \frac{\tilde{x}_{ij}}{g_2^t(\alpha_{1i})}$$

$$\frac{\partial c_i}{\partial \beta_j} = (\alpha_{0i} - \alpha_{1i}) \frac{x_{ij}}{g_3^t(\gamma_i)}$$

$$\frac{\partial c_{i1}}{\partial a_1} = -(1-\gamma_i) \alpha_{0i} \exp(-g_1(\alpha_{0i}))$$

$$\frac{\partial c_{i2}}{\partial a_2} = \alpha_{1i} \gamma_i \exp(-g_2(\alpha_{1i})) \frac{\partial c_{i3}}{\partial a_3} = \gamma_i (\alpha_{0i} -$$

$$\alpha_{1i}) \exp(-g_3(\gamma_i))$$

Para el caso del mínimo

$$\frac{\partial c_{i1}}{\partial a_1} = -(1 - \gamma_i)(1 - \alpha_{0i}) \exp(g_1(\alpha_{0i})) \quad \frac{\partial c_{i2}}{\partial a_2} = (1 - \alpha_{1i})\gamma_i \exp(g_2(\alpha_{1i}))$$

$$\frac{\partial c_{i3}}{\partial a_3} = (1 - \gamma_i)(\alpha_{0i} - \alpha_{1i}) \exp(g_3(\gamma_i))$$

Agrupando las expresiones de las derivadas parciales para un mejor cálculo:

$$\frac{\partial \left[\frac{(1 - \alpha_{0i})(Y_i^* - \mu_i^*)}{c_i^2} \right]}{\partial a_1} = \frac{-[\alpha_{0i}^t(Y_i^* - \mu_i^*) + (Y_i^* - \mu_i^*)^t(1 - \alpha_{0i})]c_i - 2c_i^t(1 - \alpha_{0i})(Y_i^* - \mu_i^*)}{c_i^3}$$

$$* \frac{\partial \left[\frac{(1 - \alpha_{1i})(Y_i^* - \mu_i^*)}{c_i^2} \right]}{\partial a_1} = \frac{(1 - \alpha_{1i})[c_i(Y_i^* - \mu_i^*)^t - 2c_i^t(Y_i^* - \mu_i^*)^t]}{c_i^3}$$

$$\frac{\partial \left[\frac{(1 - \alpha_{1i})(Y_i^* - \mu_i^*)}{c_i^2} \right]}{\partial a_2} = \frac{-[\alpha_{1i}^t(Y_i^* - \mu_i^*) + (Y_i^* - \mu_i^*)^t(1 - \alpha_{1i})]c_i - 2c_i^t(1 - \alpha_{1i})(Y_i^* - \mu_i^*)}{c_i^3}$$

$$* \frac{\partial \left[\frac{(1 - \alpha_{0i})(Y_i^* - \mu_i^*)}{c_i^2} \right]}{\partial a_2} = \frac{(1 - \alpha_{0i})[c_i(Y_i^* - \mu_i^*)^t - 2c_i^t(Y_i^* - \mu_i^*)^t]}{c_i^3}$$

Tomando la convención de que cualquier suma con límites indebidos es cero, obtenemos las derivadas de primer orden.

$$\frac{\partial(Y_i^* - \mu_i^*)}{\partial a_1} = \phi \frac{\partial \mu_i}{\partial a_1} [\psi^t(\mu_i \phi) + \psi^t(1 - \mu_i \phi)] = -\phi \frac{\partial \mu_i}{\partial a_1} H$$

$$\frac{\partial(Y_i^* - \mu_i^*)}{\partial a_2} = \phi \frac{\partial \mu_i}{\partial a_2} [\psi^t(\mu_i \phi) + \psi^t(1 - \mu_i \phi)] = -\phi \frac{\partial \mu_i}{\partial a_2} H$$

$$\frac{\partial(Y_i^* - \mu_i^*)}{\partial a_3} = \phi \frac{\partial \mu_i}{\partial a_3} [\psi^t(\mu_i \phi) + \psi^t(1 - \mu_i \phi)] = -\phi \frac{\partial \mu_i}{\partial a_3} H$$

$$\begin{aligned}
\frac{\partial K(\theta)}{\partial \omega_j} &= \sum_{i=1}^{n_0} \frac{1}{g_{a_1}^t(\alpha_{0i})\alpha_{0i}} \tilde{x}_{ij} - \sum_{i=m+1}^n \frac{1-y_i}{c_i g_{a_1}^t(\alpha_{0i})} \tilde{x}_{ij} \\
&\quad + \phi \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)(1-\alpha_{1i})(Y_i^* - \mu_i^*)}{c_i^2 g_{a_1}^t(\alpha_{0i})} \tilde{x}_{ij} \\
\frac{\partial K(\theta)}{\partial \eta_j} &= \sum_{i=n_0+1}^m \frac{1}{g_{a_2}^t(\alpha_{0i})\alpha_{1i}} \tilde{x}_{ij} - \sum_{i=m+1}^n \frac{\gamma_i}{c_i g_{a_2}^t(\alpha_{1i})} \tilde{x}_{ij} \\
&\quad - \phi \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i})(Y_i^* - \mu_i^*)}{c_i^2 g_{a_2}^t(\alpha_{0i})} \tilde{x}_{ij} \\
\frac{\partial K(\theta)}{\partial \beta_j} &= - \sum_{i=1}^{n_0} \frac{1}{(1-\gamma_i)g_{a_3}^t(\gamma_i)} x_{ij} + \sum_{i=n_0+1}^m \frac{1}{\gamma_i g_{a_3}^t(\gamma_i)} x_{ij} + \sum_{i=m+1}^n \frac{(\alpha_{0i} - \alpha_{1i})}{c_i g_{a_3}^t(\gamma_i)} x_{ij} \\
&\quad + \phi \sum_{i=m+1}^n \frac{\gamma_i(1-\alpha_{1i})(1-\alpha_{0i})(Y_i^* - \mu_i^*)}{c_i^2 g_{a_3}^t(\gamma_i)} x_{ij} \\
\frac{\partial K(\theta)}{\partial \phi} &= \sum_{i=m+1}^n \mu_i(Y_i^* - \mu_i^*) + \log(1-\gamma_i) - \psi((1-\mu_i)\phi) + \psi(\phi) \\
\frac{\partial K(\theta)}{\partial a_1} &= \sum_{i=1}^{n_0} \frac{\alpha_{0i}^t}{\alpha_{0i}} + \sum_{i=m+1}^n \frac{c_{i1}^t}{c_{i1}} + \phi \sum_{i=m+1}^n (Y_i^* - \mu_i^*) \frac{\partial \mu_i}{\partial a_1} \\
\frac{\partial K(\theta)}{\partial a_2} &= \sum_{i=1}^{n_0} \frac{\alpha_{1i}^t}{\alpha_{1i}} + \sum_{i=m+1}^n \frac{c_{i2}^t}{c_{i2}} + \phi \sum_{i=m+1}^n (Y_i^* - \mu_i^*) \frac{\partial \mu_i}{\partial a_2} \\
\frac{\partial K(\theta)}{\partial a_3} &= - \sum_{i=1+n_0}^m \frac{\gamma_{i3}^t}{(1-\gamma_i)} + \sum_{i=n_0+1}^m \frac{\gamma_{i3}^t}{\gamma} + \sum_{i=m+1}^n \frac{c_{i3}^t}{c} + \phi \sum_{i=m+1}^n (Y_i^* - \mu_i^*) \frac{\partial \mu_i}{\partial a_3}
\end{aligned}$$

Calculando las segundas derivadas:

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \omega_h \omega_j} &= - \sum_{i=1}^{n_0} \frac{g'_1(\alpha_{0i}) + \alpha_{0i} g''_1(\alpha_{0i})}{\alpha_{0i}^2 g'_1(\alpha_{0i})^3} \bar{x}_{ih} \bar{x}_{ij} + \sum_{i=m+1}^n \frac{(c_i g''_1(\alpha_{0i}) - (1 - \gamma_i) g'_1(\alpha_{0i})) (1 - \gamma_i)}{c_i^2 g'_1(\alpha_{0i})^3} \bar{x}_{ih} \bar{x}_{ij} \\
&\quad - \phi^2 \sum_{i=m+1}^n \frac{\gamma_i^2 (1 - \gamma_i)^2 (1 - \alpha_{1i})^2 H_i}{c_i^4 g'_1(\alpha_{0i})^2} \bar{x}_{ih} \bar{x}_{ij} \\
&\quad - \phi \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i) (1 - \alpha_{1i}) (c_i g''_1(\alpha_{0i}) - 2(1 - \gamma_i) g'_1(\alpha_{0i})) (\gamma_i^* - \mu_i^*)}{c_i^3 g'_1(\alpha_{0i})^3} \bar{x}_{ih} \bar{x}_{ij},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \eta_h \eta_j} &= - \sum_{i=n_0+1}^m \frac{g'_2(\alpha_{1i}) + \alpha_{1i} g''_2(\alpha_{1i})}{\alpha_{1i}^2 g'_2(\alpha_{1i})^3} \check{x}_{ih} \check{x}_{ij} + \sum_{i=m+1}^n \frac{(c_i g''_2(\alpha_{1i}) - \gamma_i g'_2(\alpha_{1i})) \gamma_i}{c_i^2 g'_2(\alpha_{1i})^3} \check{x}_{ih} \check{x}_{ij} \\
&\quad - \phi^2 \sum_{i=m+1}^n \frac{\gamma_i^2 (1 - \gamma_i)^2 (1 - \alpha_{0i})^2 H_i}{c_i^4 g'_2(\alpha_{1i})^2} \check{x}_{ih} \check{x}_{ij} \\
&\quad + \phi \sum_{i=m+1}^n \frac{\gamma_i (1 - \gamma_i) (1 - \alpha_{0i}) (c_i g''_2(\alpha_{1i}) - 2\gamma_i g'_2(\alpha_{1i})) (\gamma_i^* - \mu_i^*)}{c_i^3 g'_2(\alpha_{1i})^3} \check{x}_{ih} \check{x}_{ij},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \beta_h \beta_j} &= \sum_{i=1}^{n_0} \frac{(1 - \gamma_i) g''_3(\gamma_i) - g'_3(\gamma_i)}{(1 - \gamma_i)^2 g'_3(\gamma_i)^3} x_{ih} x_{ij} - \sum_{i=n_0+1}^m \frac{g'_3(\gamma_i) + \gamma_i g''_3(\gamma_i)}{\gamma_i^2 g'_3(\gamma_i)^3} x_{ih} x_{ij} \\
&\quad - \sum_{i=m+1}^n \frac{(\alpha_{0i} - \alpha_{1i}) (c_i g''_3(\gamma_i) + (\alpha_{0i} - \alpha_{1i}) g'_3(\gamma_i))}{c_i^2 g'_3(\gamma_i)^3} x_{ih} x_{ij} \\
&\quad - \phi^2 \sum_{i=m+1}^n \frac{H_i (1 - \alpha_{0i})^2 (1 - \alpha_{1i})^2}{c_i^4 g'_3(\gamma_i)^2} x_{ih} x_{ij} \\
&\quad - \phi \sum_{i=m+1}^n \frac{(1 - \alpha_{0i}) (1 - \alpha_{1i}) (c_i g''_3(\gamma_i) + 2(\alpha_{0i} - \alpha_{1i}) g'_3(\gamma_i)) (\gamma_i^* - \mu_i^*)}{c_i^3 g'_3(\gamma_i)^3} x_{ih} x_{ij},
\end{aligned}$$

$$\frac{\partial^2 K(\theta)}{\partial \phi^2} = (n - m) \Psi'(\phi) - \sum_{i=m+1}^n (\mu_i^2 \Psi'(\mu_i \phi) + (1 - \mu_i)^2 \Psi'((1 - \mu_i) \phi)),$$

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \omega_h \eta_j} &= - \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)}{c_i^2 g_1'(\alpha_{0i}) g_2'(\alpha_{1i})} \tilde{x}_{ih} \tilde{x}_{ij} \\
&\quad - \phi \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i} + \gamma_i(\alpha_{0i} + \alpha_{1i} - 2))(y_i^* - \mu_i^*)}{c_i^3 g_1'(\alpha_{0i}) g_2'(\alpha_{1i})} \tilde{x}_{ih} \tilde{x}_{ij} \\
&\quad + \phi^2 \sum_{i=m+1}^n \frac{(1-\alpha_{0i})(1-\alpha_{1i})\gamma_i^2(1-\gamma_i)^2 H_i}{c_i^4 g_1'(\alpha_{0i}) g_2'(\alpha_{1i})} \tilde{x}_{ih} \tilde{x}_{ij}, \\
&\quad \dots
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \omega_h \beta_j} &= \sum_{i=m+1}^n \frac{1-\alpha_{1i}}{c_i^2 g_1'(\alpha_{0i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij} \\
&\quad + \phi \sum_{i=m+1}^n \frac{(1-\alpha_{1i})(1-\alpha_{0i} + \gamma_i(\alpha_{0i} + \alpha_{1i} - 2))(y_i^* - \mu_i^*)}{c_i^3 g_1'(\alpha_{0i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij} \\
&\quad - \phi^2 \sum_{i=m+1}^n \frac{(1-\alpha_{0i})(1-\alpha_{1i})^2 \gamma_i(1-\gamma_i) H_i}{c_i^4 g_1'(\alpha_{0i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij},
\end{aligned}$$

$$\frac{\partial^2 K(\theta)}{\partial \omega_j \partial \phi} = \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)(1-\alpha_{1i})(y_i^* - \mu_i^* - \phi \mu_i \Psi'(\mu_i \phi) + \phi(1-\mu_i)\Psi'((1-\mu_i)\phi))}{c_i^2 g_1'(\alpha_{0i})} \tilde{x}_{ij}$$

$$\begin{aligned}
\frac{\partial^2 K(\theta)}{\partial \eta_h \beta_j} &= - \sum_{i=m+1}^n \frac{1-\alpha_{0i}}{c_i^2 g_2'(\alpha_{1i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij} \\
&\quad - \phi \sum_{i=m+1}^n \frac{(1-\alpha_{0i})(1-\alpha_{0i} + \gamma_i(\alpha_{0i} + \alpha_{1i} - 2))(y_i^* - \mu_i^*)}{c_i^3 g_2'(\alpha_{1i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij} \\
&\quad + \phi^2 \sum_{i=m+1}^n \frac{(1-\alpha_{0i})^2(1-\alpha_{1i})\gamma_i(1-\gamma_i) H_i}{c_i^4 g_2'(\alpha_{1i}) g_3'(\gamma_i)} \tilde{x}_{ih} x_{ij},
\end{aligned}$$

$$\frac{\partial^2 K(\theta)}{\partial \eta_j \partial \phi} = - \sum_{i=m+1}^n \frac{\gamma_i(1-\gamma_i)(1-\alpha_{0i}) (y_i^* - \mu_i^* - \phi \mu_i \Psi'(\mu_i \phi) + \phi(1-\mu_i) \Psi'((1-\mu_i)\phi))}{c_i^2 g_2'(\alpha_{1i})} \tilde{x}_{ij}$$

$$\frac{\partial^2 K(\theta)}{\partial \beta_j \partial \phi} = \sum_{i=m+1}^n \frac{(1-\alpha_{0i})(1-\alpha_{1i}) (y_i^* - \mu_i^* - \phi \mu_i \Psi'(\mu_i \phi) + \phi(1-\mu_i) \Psi'((1-\mu_i)\phi))}{c_i^2 g_3'(\gamma_i)} x_{ij}.$$

$$\frac{\partial K(\theta)}{\delta \omega_j \delta a_1} = - \sum_{i=1}^{n_0} \frac{\alpha_{0i}^t}{g_{a_1}^t(\alpha_{0i}) \alpha_{0i}^2} \tilde{x}_{ij} + \sum_{i=m+1}^n \frac{c_i(y_i-1)}{c_i^2 g_{a_1}^t(\alpha_{0i})} \tilde{x}_{ij} + \sum_{i=m+1}^n \frac{[-\phi H \frac{\delta \mu_i}{\delta a_1}] c_i + 2c_i^2 (Y_i^* - \mu_i^*)}{c_i^3}$$

$$\frac{\partial K(\theta)}{\delta \omega_j \delta a_2} = - \sum_{i=m+1}^{n_0} \frac{c_i^t (y_i-1)}{c_i^2 g_{a_1}(\alpha_{0i})} \tilde{x}_{ij} + \sum_{i=m+1}^n \frac{\alpha_{1i}^t (Y_i^* - \mu_i^*)}{c_i^3} - \frac{[\phi H \frac{\delta \mu_i}{\delta a_1}] (1-\alpha_{1i})}{c_i^3}$$

$$\frac{\partial K(\theta)}{\delta \omega_j \delta a_3} = - \sum_{i=m+1}^n \frac{\gamma_i^t - 2c_i^t}{c_i} + \phi \sum_{i=m+1}^n [\gamma_i^t (1-2\gamma_i)(Y_i^* - \mu_i^*) - \phi \frac{\delta \mu_i}{\delta a_3} H \gamma_i (1-\gamma_i)] c_i - 2c_i^t [\gamma_i (1-\gamma_i)(Y_i^* - \mu_i^*)] \frac{1}{c_i^3}$$

$$\frac{\partial K(\theta)}{\delta \eta_j \delta a_3} = - \sum_{i=\eta_0+1}^m \frac{\alpha_{1i}^t}{\alpha_{1i} g_2^t(\alpha_{1i})} \tilde{x}_{ij} + \phi \sum_{i=m+1}^n \left[\frac{\gamma_i (1-\gamma_i) \tilde{x}_{ij}}{g_2^t(\alpha_{1i})} * \frac{\delta(1-\alpha_{0i})(Y_i^* - \mu_i^*)}{\delta a_3 c_i^2} \right]$$

9.2 Tablas y gráficos de la simulación del modelo

Tabla 16: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=0.5 caso 0-1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a1	0.50000	1.50000		0.50365	0.50000	- 0.00365	0.50000	0.01486
a2	0.50000	1.50000		0.48132	0.50000	0.01868	0.50000	0.10531
a3	0.50000	0.50000		0.46640	0.50000	0.03360	0.50000	0.20705
phi	50.00000	24.58500		27.80400	50.00000	22.19600	50.00000	23.84700
b1	0.10000	- 0.32712	- 0.22711	- 0.13417	0.32711	0.23417	0.33137	0.24262
b2	0.20000	0.79036	0.31499	0.30280	- 0.11499	- 0.10280	0.23146	0.21850
b3	- 0.10000	0.10774	- 0.12544	- 0.11506	0.02544	0.01506	0.12140	0.10350
b4	0.40000	0.27604	0.56965	0.32055	- 0.16965	0.07945	0.18622	0.10039
b5	0.10000	- 0.20707	0.13693	- 0.16534	- 0.03693	0.26534	0.11488	0.17697
eta1	0.10000	0.38187		0.58501	0.10000	- 0.48501	0.10000	0.52310
eta2	- 0.10000	- 1.06920		- 0.29516	- 0.10000	0.19516	0.10000	0.42252
eta3	- 0.20000	- 0.37595		- 0.30798	- 0.20000	0.10798	0.20000	0.23246
eta4	0.30000	0.02408		0.53545	0.30000	- 0.23545	0.30000	0.28573
eta5	0.50000	0.77727		0.46945	0.50000	0.03055	0.50000	0.19519
w1	0.50000	0.17114		0.18069	0.50000	0.31931	0.50000	0.33983
w2	- 1.00000	- 0.18460		- 0.35934	- 1.00000	- 0.64066	1.00000	0.74041
w3	- 0.50000	- 0.46278		- 0.36726	- 0.50000	- 0.13274	0.50000	0.25072
w4	0.50000	0.94846		0.86367	0.50000	- 0.36367	0.50000	0.39564
w5	1.00000	0.95210		0.93175	1.00000	0.06825	1.00000	0.20862

Fuente: Elaboración propia

Tabla 17: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1 caso 0-1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a1	1.00000	0.90000		1.04900	1.00000	- 0.04900	1.00000	0.07793
a2	1.00000	0.90000		1.09450	1.00000	- 0.09450	1.00000	0.22972
a3	1.00000	0.90000		1.22150	1.00000	- 0.22150	1.00000	1.99400
phi	200.00000	210.21000		115.54000	200.00000	84.46000	200.00000	137.01000
b1	0.10000	- 0.33519	0.84458	- 0.35975	- 0.74458	0.45975	0.74773	0.48740
b2	0.20000	0.72710	0.36348	0.29999	- 0.16348	- 0.09999	0.28160	0.12754
b3	- 0.10000	0.05038	- 0.15932	- 0.00788	0.05932	- 0.09212	0.15363	0.12551
b4	0.40000	0.33373	0.65149	0.49176	- 0.25149	- 0.09176	0.26562	0.16860
b5	0.10000	- 0.28933	0.18200	- 0.01389	- 0.08200	0.11389	0.15396	0.12265
eta1	0.10000	0.42656		- 0.11194	0.10000	0.21194	0.10000	1.67020
eta2	- 0.10000	- 1.22350		- 0.02883	- 0.10000	- 0.07117	0.10000	3.37390
eta3	- 0.20000	- 0.26893		- 0.74606	- 0.20000	0.54606	0.20000	3.36180
eta4	0.30000	- 0.23154		1.06150	0.30000	- 0.76150	0.30000	4.09860
eta5	0.50000	0.80417		0.84506	0.50000	- 0.34506	0.50000	1.38980
w1	0.50000	0.50067		0.41511	0.50000	0.08489	0.50000	0.20372
w2	- 1.00000	- 0.56741		- 0.74935	- 1.00000	- 0.25065	1.00000	0.41523
w3	- 0.50000	- 0.61081		- 0.48952	- 0.50000	- 0.01048	0.50000	0.23095
w4	0.50000	0.98824		0.79419	0.50000	- 0.29419	0.50000	0.35961
w5	1.00000	1.05090		0.99377	1.00000	0.00623	1.00000	0.22313

Fuente: Elaboración propia

Tabla 18: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1.5 caso 0-1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a1	1.50000	1.00000		1.01520	1.50000	0.48480	1.50000	0.48750
a2	1.50000	1.00000		1.20610	1.50000	0.29390	1.50000	0.36988
a3	1.50000	1.00000		1.36170	1.50000	0.13830	1.50000	1.43920
phi	200.00000	973.49000		196.73000	200.00000	3.27000	200.00000	326.24000
b1	0.10000	- 0.21034	1.58680	- 0.27347	- 1.48680	0.37347	1.48940	0.43315
b2	0.20000	0.69072	0.27050	0.21913	- 0.07050	- 0.01913	0.27598	0.27418
b3	- 0.10000	0.05974	- 0.22054	- 0.06898	0.12054	- 0.03102	0.20095	0.17843
b4	0.40000	0.31359	0.71938	0.42487	- 0.31938	- 0.02487	0.33222	0.14851
b5	0.10000	- 0.15099	0.18908	- 0.14364	- 0.08908	0.24364	0.15329	0.18780
eta1	0.10000	0.25326		0.16175	0.10000	- 0.06175	0.10000	0.74235
eta2	- 0.10000	- 1.12160		- 0.84892	- 0.10000	0.74892	0.10000	1.30750
eta3	- 0.20000	- 0.20998		- 0.31413	- 0.20000	0.11413	0.20000	0.67174
eta4	0.30000	- 0.29137		- 0.05430	0.30000	0.35430	0.30000	1.13210
eta5	0.50000	0.53538		0.60836	0.50000	- 0.10836	0.50000	0.61652
w1	0.50000	0.80147		0.63004	0.50000	- 0.13004	0.50000	0.28439
w2	- 1.00000	- 0.83368		- 0.99957	- 1.00000	- 0.00043	1.00000	0.36537
w3	- 0.50000	- 0.64273		- 0.51526	- 0.50000	0.01526	0.50000	0.24306
w4	0.50000	0.92643		0.76199	0.50000	- 0.26199	0.50000	0.35932
w5	1.00000	1.25170		1.34390	1.00000	- 0.34390	1.00000	0.44186

Fuente: Elaboración propia

Tabla 19: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=0.5 caso 0 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	0.50000	1.50000		0.57815	0.50000	- 0.07815	0.50000	0.10884
a3	0.50000	1.50000		0.21446	0.50000	0.28554	0.50000	0.30487
phi	20.00000	2.09270		2.25040	20.00000	17.74960	20.00000	17.75000
b1	0.10000	- 0.00717	- 0.53816	- 0.15016	0.63816	0.25016	0.63931	0.26837
b2	0.20000	0.08587	0.16872	0.10283	0.03128	0.09717	0.16481	0.16708
b3	- 0.10000	- 0.06142	- 0.02567	- 0.03679	- 0.07433	- 0.06322	0.11421	0.10138
b4	0.40000	0.17294	0.20040	0.21426	0.19960	0.18574	0.20846	0.19149
b5	0.10000	- 0.00316	- 0.09652	0.05842	0.19652	0.04158	0.21419	0.15388
w1	0.10000	1.43130		- 0.44303	0.10000	0.54303	0.10000	0.57550
w2	- 0.10000	0.13843		- 0.20640	- 0.10000	0.10640	0.10000	0.19845
w3	- 0.20000	0.07837		- 0.09499	- 0.20000	- 0.10501	0.20000	0.14770
w4	0.30000	0.00204		- 0.08566	0.30000	0.38566	0.30000	0.39117
w5	0.50000	- 0.14127		0.10699	0.50000	0.39301	0.50000	0.42097

Fuente: Elaboración propia

Tabla 20: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1 caso 0 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	1.00000	1.50000	-	0.47176	1.00000	0.52824	0.50000	0.05905
a3	1.00000	1.50000	-	0.24273	1.00000	0.75727	1.00000	0.76940
phi	20.00000	2.03440	-	2.26580	20.00000	17.73420	20.00000	17.73500
b1	0.10000	- 0.00588	- 0.54201	0.01191	0.64201	0.08809	0.64341	0.13809
b2	0.20000	0.07182	0.14173	0.16904	0.05827	0.03096	0.19076	0.15176
b3	- 0.10000	- 0.04878	- 0.02173	- 0.07050	- 0.07827	- 0.02950	0.13558	0.11563
b4	0.40000	0.16849	0.20536	0.18564	0.19464	0.21436	0.20106	0.24248
b5	0.10000	0.04807	- 0.07538	0.10050	0.17538	- 0.00050	0.19395	0.15423
w1	0.10000	1.42060	-	- 0.08253	0.10000	0.18253	0.10000	0.29713
w2	- 0.10000	0.11108	-	- 0.06735	- 0.10000	- 0.03266	0.10000	0.13677
w3	- 0.20000	0.06104	-	0.00758	- 0.20000	- 0.20758	0.20000	0.23491
w4	0.30000	0.01081	-	- 0.04837	0.30000	0.34837	0.30000	0.35400
w5	0.50000	- 0.16260	-	0.04710	0.50000	0.45290	0.50000	0.48195

Fuente: Elaboración propia

Tabla 21: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1.5 caso 0 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	1.50000	0.50000		0.45659	1.50000	1.04341	0.50000	0.08617
a3	1.50000	0.50000		0.30136	1.50000	1.19864	1.50000	1.20590
phi	20.00000	2.08480		2.27420	20.00000	17.72580	20.00000	17.72600
b1	0.10000	- 0.01679	- 0.54691	0.05112	0.64691	0.04888	0.64813	0.15707
b2	0.20000	0.11644	0.22834	0.12378	- 0.02834	0.07622	0.16404	0.14779
b3	- 0.10000	- 0.04344	- 0.00397	- 0.02707	- 0.09603	- 0.07293	0.13376	0.11204
b4	0.40000	0.19080	0.20804	0.16404	0.19196	0.23596	0.19842	0.25827
b5	0.10000	- 0.01509	- 0.10704	- 0.00122	0.20704	0.10122	0.22111	0.18304
w1	0.10000	1.43250		- 0.17076	0.10000	0.27076	0.10000	0.34873
w2	- 0.10000	0.12556		- 0.17457	- 0.10000	0.07457	0.10000	0.19642
w3	- 0.20000	0.06116		- 0.03507	- 0.20000	- 0.16493	0.20000	0.19878
w4	0.30000	- 0.00528		- 0.03310	0.30000	0.33310	0.30000	0.33844
w5	0.50000	- 0.13355		0.12099	0.50000	0.37901	0.50000	0.40070

Fuente: Elaboración propia

Tabla 22: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=0.5 caso 1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	0.50000	1.50000	-	0.50000	0.50000	-	0.50000	0.00002
a3	0.50000	1.50000	-	0.55470	0.50000	- 0.05470	0.50000	0.11058
phi	100.00000	1.24410	-	1.05850	100.00000	98.94150	100.00000	98.94200
b1	- 0.50000	0.77787	1.06110	0.78981	- 1.56110	- 1.28981	1.56200	1.29290
b2	1.00000	0.74985	0.63549	0.75777	0.36451	0.24223	0.41708	0.36092
b3	- 0.50000	- 0.07091	- 0.29658	- 0.07783	- 0.20342	- 0.42217	0.23689	0.44862
b4	0.50000	0.12358	0.31734	0.11674	0.18266	0.38326	0.19361	0.39071
b5	- 0.50000	- 0.43465	- 0.32475	- 0.44157	- 0.17525	- 0.05843	0.20076	0.13498
eta1	0.50000	0.59517	-	0.58328	0.50000	- 0.08328	0.50000	0.10959
eta2	- 1.00000	1.03240	-	1.02050	- 1.00000	- 2.02050	1.00000	2.03990
eta3	- 0.50000	- 0.61714	-	- 0.60540	- 0.50000	0.10540	0.50000	0.19717
eta4	0.50000	0.62880	-	0.63673	0.50000	- 0.13673	0.50000	0.16708
eta5	1.00000	- 0.48288	-	- 0.47456	1.00000	1.47456	1.00000	1.48030

Fuente: Elaboración propia

Tabla 23: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1 caso 1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	1.00000	0.50000	-	0.49912	1.00000	0.50088	1.00000	0.50111
a3	1.00000	0.50000	-	0.81192	1.00000	0.18808	1.00000	0.51213
phi	50.00000	2.15150	-	2.05690	50.00000	47.94310	50.00000	48.01500
b1	- 0.50000	- 0.13592	0.63772	0.05645	- 1.13772	- 0.55645	1.13880	0.63279
b2	1.00000	0.52661	0.46265	0.64084	0.53735	0.35916	0.56764	0.47329
b3	- 0.50000	- 0.14410	- 0.27878	- 0.26142	- 0.22122	- 0.23858	0.24517	0.34903
b4	0.50000	0.10931	0.23936	0.05586	0.26064	0.44414	0.26729	0.48265
b5	- 0.50000	- 0.30037	- 0.22010	- 0.39258	- 0.27990	- 0.10742	0.29234	0.22789
eta1	0.50000	- 1.82330	-	- 1.95650	0.50000	2.45650	0.50000	2.47070
eta2	- 1.00000	1.14550	-	1.02940	- 1.00000	- 2.02940	1.00000	2.06870
eta3	- 0.50000	- 0.77465	-	- 0.63875	- 0.50000	0.13875	0.50000	0.35339
eta4	0.50000	0.78794	-	0.85313	0.50000	- 0.35313	0.50000	0.41536
eta5	1.00000	- 0.49830	-	- 0.39958	1.00000	1.39958	1.00000	1.42080

Fuente: Elaboración propia

Tabla 24: Comparación de Bias y Error Cuadrático Medio para Metodología de Enlaces Asimétricos y Papke and Wooldridge n=1000 y a=1.5 caso 1 inflacionado

Parámetros	Valor Real	Valor Inicial	mPW	mBlm	Bias(PW)	Bias(Blm)	RSME(PW)	RSME(Blm)
a2	1.50000	0.90000	-	0.99993	1.50000	0.50007	1.00000	0.00023
a3	1.50000	0.90000	-	1.02620	1.50000	0.47380	1.00000	0.06492
phi	50.00000	0.65891	-	0.66468	50.00000	49.33532	50.00000	49.33500
b1	0.10000	- 0.04699	0.93969	- 0.02827	- 0.83969	0.12827	0.84122	0.14719
b2	0.20000	0.35322	0.14101	0.36769	0.05899	- 0.16769	0.19382	0.30800
b3	- 0.10000	0.05755	- 0.07073	0.05571	- 0.02927	- 0.15571	0.11563	0.21950
b4	0.40000	0.13209	0.28418	0.44913	0.11582	- 0.04913	0.13436	0.11784
b5	0.10000	- 0.15616	0.07392	- 0.14359	0.02608	0.24359	0.10645	0.27792
eta1	0.50000	- 0.91479	-	- 0.93314	0.50000	1.43314	0.50000	1.43590
eta2	- 1.00000	0.13318	-	0.11566	- 1.00000	- 1.11566	1.00000	1.14900
eta3	- 0.50000	- 0.20234	-	- 0.18467	- 0.50000	- 0.31533	0.50000	0.36078
eta4	0.50000	0.60911	-	0.59167	0.50000	- 0.09167	0.50000	0.15123
eta5	1.00000	0.28624	-	0.27000	1.00000	0.73000	1.00000	0.74737

Fuente: Elaboración propia

9.3 Código en Matlab LOGLOG:

9.3.1 Función acumulada y sus derivadas

```
function [g0, g1, g2] = gloglog(y,a)
    g0 = -log(-(1./a)*log(y));
    g1 = -1./(y.*log(y));
    %g2 = 1./((y.^2).*log(y)) + 1./((y.^2).*log(y).^2);
    g2=(log(y)+1)./(y.*log(y).^2);
```

9.3.2 Simulación de distribuciones caso 0 inflacionado

```
if model >= 0 % initial estimation of phi and beta
    Z = log(-(1./thetav(1).*log(1-Yc)));
    Yc(Yc<=0.0000000001) = 0.0000000001;
    %Xc= X(indc,:);
    %Ym = Y+2;
    [beta1,BINT,R,RINT,STATS] = regress(Z,Xc);
    %beta0 = inv(Xc'*Xc)*Xc'*Z;
    %ee = Z - Xc*beta0;
    %sigma2 = (ee'*ee)/(m-kk);
    %mur = 1./(1 + exp(-Xc*beta0)); logistica
    mur = 1-exp(-thetav(1)*exp(-Xc*beta1)); %caso cloglog
    mur(mur<=0.0000000001) = 0.0000000001;
    mur(mur>=0.9999999999) = 0.9999999999;
    sigma2 = STATS(end);
    phi0x = 1./(sigma2*mur.*(1-mur));
    phi0 = mean(phi0x(phi0x < 1e+4)) - 1;
end
if model ~= 0 % multinomial setting
    %Ym(Ym==3)=1;
    w00 = mnrfits(X(:,2:end),Ym,'model','nominal','interactions','on');
end
p0model = mnrfits(w00, X(:,2:end));
%mur2 = p0model(indc,2) + p0model(indc,3).*mur;
mur2 = p0model(indc,2) + (1 - p0model(indc,1) - p0model(indc,2)).*mur;
alpha00 = p0model(indc,1)./(1-mur2);
alpha01 = p0model(indc,2)./mur2;
w1 = regress(log(-(1./thetav(1).*log(1-alpha00))),Xc);
eta1 = regress(log(-(1./thetav(1).*log(1-alpha01))),Xc);
9.3.3 Simulación de distribuciones caso 0-1 inflacionado
if model >= 0 % initial estimation of phi and beta
```

```

Z = log(-(1./thetav(1).*log(1-Yc)));
Yc(Yc<=0.0000000001) = 0.0000000001;
%Xc= X(indc,:);
%Ym = Y+2;
[beta1,BINT,R,RINT,STATS] = regress(Z,Xc);
%beta0 = inv(Xc'*Xc)*Xc'*Z;
%ee = Z - Xc*beta0;
%sigma2 = (ee'*ee)/(m-kk);
%mur = 1./(1 + exp(-Xc*beta0)); logistica
mur = 1-exp(-thetav(1)*exp(-Xc*beta1)); %caso cloglog
mur(mur<=0.0000000001) = 0.0000000001;
mur(mur>=0.9999999999) = 0.9999999999;
sigma2 = STATS(end);
phi0x = 1./(sigma2*mur.*(1-mur));
phi0 = mean(phi0x(phi0x < 1e+4)) - 1;
end
if model ~= 0 % multinomial setting
    %Ym(Ym==3)=1;
    w00 = mnrfits(X(:,2:end),Ym,'model','nominal','interactions','on');
end
p0model = mnrfits(w00, X(:,2:end));
%mur2 = p0model(indc,2) + p0model(indc,3).*mur;
mur2 = p0model(indc,2) + (1 - p0model(indc,1) - p0model(indc,2)).*mur;
alpha00 = p0model(indc,1)./(1-mur2);
alpha01 = p0model(indc,2)./mur2;
w1 = regress(log(-(1./thetav(1).*log(1-alpha00))),Xc);
eta1 = regress(log(-(1./thetav(1).*log(1-alpha01))),Xc);

```

9.3.4 Gradiente y Matriz Hessiana para 1 inflacionado

```

function [f, grad, H] =pkfun1_loglog(thetav,X,Y)
%global res;
[n,kk] = size(X);
phi = thetav(3);
% new parameters
a2 = thetav(2);
a3 = thetav(3);
index1 = find(Y==1);
n1 = length(index1);
indc = setdiff((1:n)',index1);
% indices modificados
pred =X*[thetav(4:kk+3) thetav(kk+4:end)];

```

```

32); pred = pred.*(pred<=32).*(pred>=-32) - 32*(pred < -32) + 32*(pred >
F = 1-exp(-thetav(2)*exp(pred));
gamma = F(:,1);
gamma(gamma<=0.0001) = 0.0001;
gamma(gamma>=0.9999) = 0.9999;
alpha1 = F(:,2);
alpha1(alpha1<=0.0001) = 0.0001;
alpha1(alpha1>=0.9999) = 0.9999;
gc = gamma(indc);
gc(gc<=0.0001) = 0.0001;
gc(gc>=0.9999) = 0.9999;
c = 1 - alpha1(indc).*gc;
mu = gc.*(1- alpha1(indc))./c;
Yc = Y(indc);
aux1 = sum(log(alpha1(index1).*gamma(index1)));
bet = betapdf(Yc,phi*mu,phi*(1-mu));
hh = c.*bet;
hh(hh <= 0.3e-323) = 0.3e-323;
aux2 = sum(log(hh));
f = -(aux1+ aux2); %- loglikelihood function
if isnan(f)==1
    f = realmax;
    grad = ones(2*kk + 1,1);
    H = ones(2*kk+1,2*kk+1);
else
    if nargout > 1 % with gradient

        [g10, g11, g12] = gloglog(alpha1,a2);
        [gc0, gc1, gc2] = gloglog(gamma,a3);

        xs = [X(index1,:) ; X(indc,:)];

        g10s = [g10(index1); g10(indc)];
        gc0s = [gc0(index1); gc0(indc)];

        g11s = [g11(index1);g11(indc)];
        gc1s = [gc1(index1); gc1(indc)];

        yh = log(Yc./(1-Yc));
        muh = psi(phi*mu) - psi(phi*(1-mu));

```

```

dmua2=-gc.*(1-gc).*(-alpha1(indc)).*exp(g10(indc))./(c.^2);
dmua3=(-alpha1(indc)).*(1-gc).*exp(gc0(indc))./(c.^2);

dmua21=-gc.*(1-gc).*(-alpha1(indc))./(c.^2);%exp(g10(indc))
dmua31=(-alpha1(indc)).*(-gc)./(c.^2);%exp(gc0(indc))

% Gradient

muh)];
grada2 = [-ones(n1,1);-(-alpha1(indc)).*gc./c-dmua21.*phi.*(yh -
grada22=sum(grada2.*exp(g10s));

grada3 = [-ones(n1,1);dmua31.*phi.*(yh - muh)];
grada33=sum(grada3.*exp(gc0s));

psi(phi));
gradphi = sum(mu.*(yh - muh) + log(1-Yc) - psi(phi*(1-mu)) +
Mg = [1./gamma(index1) ; -alpha1(indc)./c + phi*(1-
alpha1(indc)).*(yh-muh)./(c.^2)];
gradg = (Xs')*(Mg./gc1s);
Mg1 = [1./alpha1(index1) ; -gc./c - (phi*gc.*(1-gc).*(yh-
muh))./(c.^2)];
grad1 = (Xs')*(Mg1./g11s);

grad = -[grada22; grada33; gradphi ; gradg ; grad1]; % gradiente

if nargout > 2 % with Hessian

Hx = psi(1, phi*mu) + psi(1,phi*(1-mu));
Hphi = (n - n1)*psi(1,phi) - sum((mu.^2).*psi(1,phi*mu)+ ((1 -
mu).^2).*psi(1,phi*(1-mu)));
auxx1 = yh-muh-phi*mu.*psi(1,phi*mu)+phi*(1-mu).*psi(1,(1-mu)*phi);
auxx2 = 1 + gc.*(alpha1(indc) -2);

%second derivate w.r.t to beta and phi
auxphib = (1-alpha1(indc)).*auxx1./(gc1(indc).*(c.^2));
Hphib = (X(indc,:))'*auxphib;

%second derivate w.r.t to beta
auxb1 = alpha1(indc).*(c.*gc2(indc) -
alpha1(indc).*gc1(indc))./((c.^2).*(gc1(indc).^3));

```

```

    auxb2 = - (phi^2)*Hx.*((1-
alpha1(indc)).^2)./((c.^4).*(gc1(indc).^2));
    auxb3 = phi*(alpha1(indc)-1).*(yh-muh).*(c.*gc2(indc) -
2*alpha1(indc).*(gc1(indc)))./((c.^3).*(gc1(indc).^3));
    auxb = [ -(gc2(index1).*(gamma(index1) +
gc1(index1)))./((gamma(index1).^2).*(gc1(index1).^3)) ; auxb1 + auxb2 +
auxb3];
    auxb = sparse(auxb);
    B = diag(auxb);
    Hb = Xs'*B*Xs;

%second derivate w.r.t to eta and phi
    auxphiw1 = - gc.*(1-gc).*auxx1./(g11(indc).*(c.^2));
    Hphiw1 = (X(indc,:))'*auxphiw1;

%second derivate w.r.t to beta and eta
    auxbw11 = -1./((c.^2).*(g11(indc).*(gc1(indc))));
    auxbw12 = -phi*(yh-muh).*auxx2./((c.^3).*(g11(indc).*(gc1(indc))));
    auxbw13 = (phi^2)*(1 - alpha1(indc)).*gc.*(1-
gc).*Hx./((c.^4).*(g11(indc).*(gc1(indc))));
    auxbw1 = auxbw11 + auxbw12 + auxbw13;
    auxbw1 = sparse(auxbw1);
    Bw1 = diag(auxbw1);
    Hbw1 = X(indc,:)'*Bw1*X(indc,:);

%second derivate w.r.t to eta
    auxw11 = (c.*g12(indc)- gc.*g11(indc)
).*(gc)./((c.^2).*(g11(indc).^3));
    auxw12 = -(phi^2)*(gc.^2).*((1-
gc).^2).*Hx./((c.^4).*(g11(indc).^2));
    auxw13 = phi*gc.*(1-gc).*(yh-muh).*(c.*g12(indc)-
2*gc.*g11(indc))./((c.^3).*(g11(indc).^3));
    auxw1 = [ -(g11(index1) +
alpha1(index1).*(g12(index1)))./((alpha1(index1).^2).*(g11(index1).^3)); auxw11
+ auxw12 + auxw13];
    auxw1 = sparse(auxw1);
    Bw1 = diag(auxw1);
    Hw1 = Xs'*Bw1*Xs;

% second derivative a2 to a2
    aux1Ha2=-sum(exp(g10(index1))./(alpha1(index1).^2));
    aux2Ha2=-sum((-alpha1(indc)).*(gamma(indc).*(-
c+gamma(indc)).*exp(2*g10(indc)))./((c.^2)));
    aux3Ha2=sum(phi.*Hx+(-alpha1(indc)).*exp(g10(indc)).*(yh-muh).*(-
c+2*(1-alpha1(indc)).*(gamma(indc)))./((c.^3)));

```



```

Ha2 = aux1Ha2+aux2Ha2+aux3Ha2;

% second derivative a3 to a3
aux1Ha3=-sum((1-
gamma(index1)).*exp(2*gc0(index1))./(gamma(index1).^2));
aux2Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*phi.*dmua3.*Hx.*(-gamma(indc))./(c.^2));
aux3Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*(-gamma(indc)).*(yh-
muh).*exp(gc0(indc))./(c.^2));
aux3Ha4=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc))*2.*((-gamma(indc)).^2).*alpha1(indc).*(yh-
muh).*exp(gc0(indc))./(c.^3));
Ha3 = aux1Ha3+aux2Ha3+aux3Ha3+aux3Ha4;

dna1=(-phi.*Hx./(c.^2)+2*(-alpha1(indc)).*gamma(indc).*(yh-
muh))./(c.^3);
%dwa1=(-phi.*Hx.*dmua1./(c.^2)+2*(1-gamma(indc)).*(yh-muh))./(c.^3);
dna2=(-phi.*Hx.*dmua2./(c.^2)+2*(-alpha1(indc)).*gamma(indc).*(yh-
muh).*exp(g10(indc)))./(c.^3);
%dwa2=-(1-alpha1(indc)).*(exp(g10(indc)).*(yh-
muh)+phi.*Hx.*dmua2)./(c.^2)+2*((1-alpha1(indc)).^2).*gamma(indc).*(yh-
muh).*exp(g10(indc))./(c.^3);

% secod derivative eta-a2
Hea2_1 = -(1-
alpha1(index1)).*exp(g10(index1))./(alpha1(index1).*g11(index1));
Hea2_2 = -(gamma(indc).^2).*(1-
alpha1(indc)).*exp(g10(indc))./((c.^2).*g11(indc));
Hea2_3 = phi.*gamma(indc).*(1-gamma(indc)).*dna2./g11(indc);
Hea2_ = [Hea2_1;Hea2_2+Hea2_3];
Hea2 = (Xs')*(Hea2_);

% secod derivative eta-a3
Hea3_1 = -(1-
gamma(indc)).*(c+alpha1(indc).*gamma(indc)).*exp(gc0(indc))./((c.^2).*g11(ind
c));
Hea3_2 = 2*phi*alpha1(indc).*((1-
gamma(indc)).^2).*gamma(indc).*(yh-muh).*exp(gc0(indc))./((c.^3).*g11(indc));
Hea3_3 = -phi*(1-gamma(indc)).*(1-2*gamma(indc)).*(yh-
muh).*exp(gc0(indc))./((c.^2).*g11(indc));
Hea3_4 = phi*phi*(1-
gamma(indc)).*gamma(indc).*Hx.*dmua3./((c.^2).*g11(indc));
Hea3_ = [zeros(n1,1) ;Hea3_1+Hea3_2+Hea3_3+Hea3_4];
Hea3 = (Xs')*(Hea3_);

```

```

% secod derivative beta-a2
Hba2_1 = (1-alpha1(indc)).*(-c+(-
alpha1(indc)).*gamma(indc)).*exp(g10(indc))./((c.^2).*gc1(indc));
Hba2_2 = phi.*dna1./gc1(indc);
Hba2_ = [zeros(n1,1);Hba2_1+Hba2_2];
Hba2 = (Xs')*(Hba2_);

% secod derivative beta-a3
Hba3_1 = -(1-
gamma(index1)).*exp(gc0(index1))./((gamma(index1).^2).*gc1(index1));
Hba3_2 = (1-
gamma(indc)).*alpha1(indc).*alpha1(indc).*exp(gc0(indc))./((c.^2).*gc1(indc))
;
Hba3_3 = phi*(1-alpha1(indc)).*(-
phi.*dmua3.*Hx.*(c.^2)+2*c.*alpha1(indc).*(1-gamma(indc)).*Hx.*(yh-
muh)).*exp(gc0(indc))./((c.^4).*gc1(indc));
Hba3_ = [Hba3_1;Hba3_2+Hba3_3];
Hba3 = (Xs')*(Hba3_);

% secod derivative phi-a2
Hpa2_1 = gamma(indc).*(1-gamma(indc)).*(1-
alpha1(indc)).*(exp(g10(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa2 = sum(Hpa2_1);

% secod derivative phi-a3
Hpa2_1 = (1-gamma(indc)).*(1-alpha1(indc)).*(exp(gc0(indc))).*((yh-
muh)-phi*Hx.*mu+phi*psi(1,phi*(1-mu)))./(c.^2);
Hpa3 = sum(Hpa2_1);

% second derivative a2a3
Ha2a3_1 = -sum((1-
gamma(indc)).*(c+alpha1(indc)).*gamma(indc)).*exp(gc0(indc))./(c.^2));
Ha2a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua2);
Ha2a3_3 = -sum((1-
gamma(indc)).*(c+2*alpha1(indc)).*gamma(indc)).*(yh-muh)./(c.^3));
Ha2a3 = Ha2a3_1+Ha2a3_2+Ha2a3_3;

%second derivate w.r.t to eta and omega
auxw011 = gc.*(gc-1)./((c.^2).*g11(indc));
auxw012 = -phi*gc.*(1-gc).*(yh-muh).*auxx2./((c.^3).*g11(indc));
auxw013 = (phi^2).*(1-alpha1(indc)).*(gc.^2).*((1-
gc).^2).*Hx./((c.^4).*g11(indc));
auxw01 = auxw011 + auxw012 + auxw013;
auxw01 = sparse(auxw01);

```

```

Bw01 = diag(auxw01);
Hbw01 = X(indc,:)'*Bw01*X(indc,:);

%Hessian matrix
%H = -[Hphi Hphib' Hphiw1'; Hphib Hb Hbw1 ; Hphiw1 Hbw1 Hw1]; %
Hessian matrix

H = -[
    Ha2    ,Ha2a3    , Hpa2    , Hba2' , Hea2' ;
    Ha2a3 , Ha3     , Hpa3    , Hba3' , Hea3' ;
    Hpa2   , Hpa3   , Hphi    , Hphib' , Hphiw1';
    Hba2   , Hba3   , Hphib   , Hb     , Hbw1 ;
    Hea2   , Hea3   , Hphiw1  , Hbw1  , Hw1   ];

end
end
end
%res = vertcat(res,[theta',f]);

```

9.3.5 Gradiente y Hessian para 0-1 Inflacionado

```

function [f, grad, H] = pkfun01_loglog(thetav,X,Y)
[n, kk] = size(X);
phi = thetav(4);
a1 = thetav(1);
a2 = thetav(2);
a3 = thetav(3);
index0 = find(Y == 0.0000);
n0 = length(index0);
index1 = find(Y == 1.0000);
n1 = length(index1);
indc = union(index0,index1);
indc = setdiff((1:n)',indc);
pred = X*[thetav(5:kk+4) thetav(kk+5:2*kk+4) thetav(2*kk+5:end)];
pred = pred.*(pred<=12).*(pred>=-700) - 700*(pred < -700) +
12*(pred > 12);
% function loglog
F = exp(-thetav(1)*exp(-pred));
% function cloglog
%F = 1-exp(-thetav(1)*exp(pred));
gamma = F(:,1);
gamma(gamma==1.0000)=0.9999;

```

```

gamma(gamma<=0.0001) = 0.0001;
alpha0 = F(:,2);
alpha0(alpha0==1.0000)=0.9999;
alpha0(alpha0<=0.0001) = 0.0001;
alpha1 = F(:,3);
alpha1(alpha1==1.0000)=0.9999;
alpha1(alpha1<=0.0001) = 0.0001;
gc = gamma(indc);
gc(gc<=0.0001) = 0.0001;
gc(gc>=0.9999) = 0.9999;
c = 1 - alpha0(indc) + (alpha0(indc) - alpha1(indc)).*gc;
mu = gc.*(1- alpha1(indc))./c;
mu(mu<=0.0001) = 0.0001;
mu(mu>=0.9999) = 0.9999;
Yc = Y(indc);
aux0 = sum(log(alpha0(index0).*(1 - gamma(index0))));
aux1 = sum(log(alpha1(index1).*gamma(index1)));
bet = betapdf(Yc,phi*mu,phi*(1-mu));
hh = c.*bet;
hh(hh <= 0.3e-323) = 0.3e-323;
aux2 = sum(log(hh));
f = -(aux0 + aux1+ aux2); %- loglikelihood function
if isnan(f)==1
    f = realmax;
    grad = ones(3*kk+4,1);
    H = ones(3*kk+4,3*kk+4);
else

if nargout > 1 % with gradient
    [g00, g01, g02] = gloglog(alpha0, a1);
    [g10, g11, g12] = gloglog(alpha1, a2);
    [gc0, gc1, gc2] = gloglog(gamma, a3);
    % dim(xs) = dim(x)
    xs = [X(index0,:) ; X(index1,:) ; X(indc,:)];
    g00s = [g00(index0); g00(index1); g00(indc)];
    g10s = [g10(index0); g10(index1); g10(indc)];
    gc0s = [gc0(index0); gc0(index1); gc0(indc)];

    g01s = [g01(index0); g01(index1); g01(indc)];
    g11s = [g11(index0); g11(index1); g11(indc)];

```

```

gc1s = [gc1(index0); gc1(index1); gc1(indc)];

yh = log(Yc./(1-Yc));

muh = psi(phi*mu) - psi(phi*(1-mu));
gradphi = sum(mu.*(yh - muh) + log(1-Yc) - psi(phi*(1-mu)) +
psi(phi));
Mg = [ 1./(gamma(index0)-1) ; 1./gamma(index1) ; (alpha0(indc)-
alpha1(indc))./c + phi*(1-alpha0(indc)).*(1-alpha1(indc)).*(yh-muh)./(c.^2)];
gradg = (Xs')*(Mg./gc1s);
Mg0 = [1./alpha0(index0) ; zeros(n1,1) ; (gc-1)./c +
(phi*gc.*(1-gc)).*(1-alpha1(indc)).*(yh-muh))./(c.^2)];
grad0 = (Xs')*(Mg0./g01s);
Mg1 = [zeros(n0,1) ; 1./alpha1(index1) ; -gc./c + (phi*gc.*(1-
gc).*(alpha0(indc)-1).*(yh-muh))./(c.^2)];
grad1 = (Xs')*(Mg1./g11s);

dmua1=gc.*(1-gc).*(1-alpha1(indc)).*(1-
alpha0(indc)).*exp(g00(indc))./(c.^2);
dmua2=-gc.*(1-gc).*(1-alpha0(indc)).*(1-
alpha1(indc)).*exp(g10(indc))./(c.^2);
dmua3=(1-alpha1(indc)).*(1-alpha0(indc)).*(1-
gc).*exp(gc0(indc))./(c.^2);

dmua11=gc.*(1-gc).*(1-alpha1(indc)).*(1-
alpha0(indc))./(c.^2);%exp(g00(indc))
dmua21=-gc.*(1-gc).*(1-alpha0(indc)).*(1-
alpha1(indc))./(c.^2);%exp(g10(indc))
dmua31=(1-alpha1(indc)).*(1-alpha0(indc)).*(1-
gc)./(c.^2);%exp(gc0(indc))

%grada1 = sum((1-alpha0(index0)).*g00(index0)./alpha0(index0))-
sum((1-alpha0(indc)).*(1-gc).*g00(indc)./c) +sum(dmua1.*phi.*(yh - muh));
grada1 = [(1-alpha0(index0))./alpha0(index0);zeros(n1,1);-(1-
alpha0(indc)).*(1-gc)./c + dmua11.*phi.*(yh - muh)];
grada11=sum(grada1.*exp(g00s));

%grada2 = sum((1-alpha1(index1)).*g01(index1)./alpha1(index1))-
sum((1-alpha1(indc)).*gc.*g01(indc)./c) + sum(dmua2.*phi.*(yh - muh));
grada2 = [zeros(n0,1);(1-alpha1(index1))./alpha1(index1);-(1-
alpha1(indc)).*gc./c-dmua21.*phi.*(yh - muh)];
grada22=sum(grada2.*exp(g10s));

%grada3 = sum(-gc0(index0))+ sum((1-
gamma(index1)).*gc0(index1))./gamma(index1) + sum(dmua3.*phi.*(yh - muh));

```

```

        grada3 = [-ones(n0,1);(1-
gamma(index1))./gamma(index1);dmua31.*phi.*(yh - muh)];
        grada33=sum(grada3.*exp(gc0s));

        grad = -[grada11; grada22; grada33; gradphi; gradg ; grad0 ;
grad1]; % gradiente
    end

    if nargout > 2 % with Hessian

        Hx = psi(1, phi*mu) + psi(1,phi*(1-mu));
        Hphi = (n - n0 - n1)*psi(1,phi) - sum((mu.^2).*psi(1,phi*mu)+
((1 - mu).^2).*psi(1,phi*(1-mu)));
        auxx1 = yh-muh-phi*mu.*psi(1,phi*mu)+phi*(1-mu).*psi(1,(1-
mu)*phi);

        %second derivate w.r.t to beta and phi
        auxphib = (1-alpha0(indc)).*(1-
alpha1(indc)).*auxx1./(gc1(indc).*(c.^2));
        Hphib = (X(indc,:))'*auxphib;

        %second derivate w.r.t to beta
        %cg = c.*gc1(indc);
        %cg = cg.*(cg<=5.6438e+102) + 5.6438e+102.*(cg > 5.6438e+102);
        auxb1 = -(alpha0(indc)-alpha1(indc)).*((alpha0(indc)-
alpha1(indc)).*gc1(indc) + c.*gc2(indc))./((c.^2).*(gc1(indc).^3));
        auxb2 = - ((phi^2)*Hx.*((1 -alpha0(indc)).^2).*((1-
alpha1(indc)).^2))./((c.^4).*(gc1(indc).^2));
        nauxb3 = -phi*(1-alpha0(indc)).*(1- alpha1(indc)).*(yh-
muh).*(c.*gc2(indc) + 2*(alpha0(indc)- alpha1(indc)).*gc1(indc));
        dauxb3 = (c.^3).*(gc1(indc).^3);
        nauxb3(nauxb3 ==Inf) = realmax;
        dauxb3(dauxb3 ==Inf) = realmax;
        auxb3 = nauxb3./dauxb3;
        auxb = [(gc2(index0).*(1-gamma(index0))- gc1(index0))./(((1 -
gamma(index0)).^2).*(gc1(index0).^3)) ; -(gc2(index1).*gamma(index1) +
gc1(index1))./((gamma(index1).^2).*(gc1(index1).^3)) ; auxb1 + auxb2 +
auxb3];

        %auxb(auxb==Inf|isnan(auxb)==1) = 0
        %[-phi*(1-alpha0(indc)).*(1- alpha1(indc)).*(yh-
muh).*(c.*gc2(indc) + 2*(alpha0(indc)- alpha1(indc)).*gc1(indc))
(c.^3).*(gc1(indc).^3) auxb3]
        auxb = sparse(auxb);
        B = diag(auxb);
        Hb = Xs'*B*Xs;

```

```

%Hb(Hb==Inf|isnan(Hb)==1) = 0

%second derivate w.r.t to omega and phi
auxphiw0 = gc.*(1-gc).*(1-
alpha1(indc)).*auxx1./(g01(indc).*(c.^2));
Hphiw0 = (X(indc,:))'*auxphiw0;

%second derivate w.r.t to eta and phi
auxphiw1 = gc.*(1-gc).*(alpha0(indc)-
1).*auxx1./(g11(indc).*(c.^2));
Hphiw1 = (X(indc,:))'*auxphiw1;

%second derivate w.r.t to beta and omega
auxx2 = 1 - alpha0(indc) + gc.*(alpha0(indc) + alpha1(indc) -2);
auxbw01 = (1-alpha1(indc))./((c.^2).*g01(indc).*gc1(indc));
auxbw02 = phi*(1 - alpha1(indc)).*(yh -
muh).*auxx2./((c.^3).*g01(indc).*gc1(indc));
auxbw03 = -(phi^2).*(1 - alpha0(indc)).*((1-
alpha1(indc)).^2).*gc.*(1- gc).*Hx./((c.^4).*g01(indc).*gc1(indc));
auxbw0 = auxbw01 + auxbw02 + auxbw03;
auxbw0 = sparse(auxbw0);
Bw0 = diag(auxbw0);
Hbw0 = X(indc,:)'*Bw0*X(indc,:);

%second derivate w.r.t to beta and eta
auxbw11 = (alpha0(indc)-1)./((c.^2).*g11(indc).*gc1(indc));
auxbw12 = -phi*(1- alpha0(indc)).*(yh-
muh).*auxx2./((c.^3).*g11(indc).*gc1(indc));
auxbw13 = (phi^2).*((1 - alpha0(indc)).^2).*(1 -
alpha1(indc)).*gc.*(1- gc).*Hx./((c.^4).*g11(indc).*gc1(indc));
auxbw1 = auxbw11 + auxbw12 + auxbw13;
auxbw1 = sparse(auxbw1);
Bw1 = diag(auxbw1);
Hbw1 = X(indc,:)'*Bw1*X(indc,:);

%second derivate w.r.t to omega
auxw01 = ((gc-1).*g01(indc) + c.*g02(indc)).*(1-
gc)./((c.^2).*g01(indc).^3));
auxw02 = -(phi^2).*(gc.^2).*((1-gc).^2).*((1-
alpha1(indc)).^2).*Hx./((c.^4).*g01(indc).^2));
auxw03 = -phi*gc.*(1 -gc).*(1 - alpha1(indc)).*(yh -
muh).*(c.*g02(indc) - 2*(1-gc).*g01(indc))./((c.^3).*g01(indc).^3));
auxw0 = [-(g01(index0) +
alpha0(index0).*g02(index0))./((alpha0(index0).^2).*g01(index0).^3));
zeros(n1,1); auxw01 + auxw02+ auxw03];

```

```

auxw0 = sparse(auxw0);
Bw0 = diag(auxw0);
Hw0 = Xs'*Bw0*Xs;

%second derivate w.r.t to eta
auxw11 = (c.*g12(indc)- gc.*g11(indc)
).*gc./((c.^2).*(g11(indc).^3));
auxw12 = -(phi^2)*(gc.^2).*((1-gc).^2).*((1-
alpha0(indc)).^2).*Hx./((c.^4).*(g11(indc).^2));
auxw13 = phi*gc.*(1-gc).*(1- alpha0(indc)).*(yh-
muh).*(c.*g12(indc)- 2*gc.*g11(indc))./(c.^3).*(g11(indc).^3));
auxw1 = [zeros(n0,1); -(g11(index1) +
alpha1(index1).*g12(index1))./(alpha1(index1).^2).*(g11(index1).^3)]; auxw11
+ auxw12 + auxw13];
auxw1 = sparse(auxw1);
Bw1 = diag(auxw1);
Hw1 = Xs'*Bw1*Xs;

```

```

%second derivate w.r.t to eta and omega

```

```

auxw011 = gc.*(gc-1)./((c.^2).*g01(indc).*g11(indc));
auxw012 = -phi*gc.*(1-gc).*(yh-
muh).*auxx2./((c.^3).*g01(indc).*g11(indc));
auxw013 = (phi^2)*(1- alpha0(indc)).*(1-
alpha1(indc)).*(gc.^2).*((1-gc).^2).*Hx./((c.^4).*g01(indc).*g11(indc));
auxw01 = auxw011 + auxw012 + auxw013;
auxw01 = sparse(auxw01);
Bw01 = diag(auxw01);
Hbw01 = X(indc,:)'*Bw01*X(indc,:);

```

```

% second derivative a1 to a1

```

```

aux1Ha1=-sum((1-
alpha0(index0)).*exp(2*g00(index0))./(alpha0(index0).^2));
aux2Ha1=sum((1-alpha0(indc)).*(1-gamma(indc)).*(c+(1-
gamma(indc)).*alpha0(indc)).*exp(2*g00(indc))./(c.^2));
aux3Ha1=-sum(phi.*Hx.*(dmua1.^2)+(1-
alpha0(indc)).*exp(g00(indc)).*(yh-muh).*(c-2*(1-alpha0(indc)).*(1-
gamma(indc)))./(c.^3));
Ha1 = aux1Ha1+aux2Ha1+aux3Ha1;

```

```

% second derivative a2 to a2

```

```

aux1Ha2=-sum((1-
alpha0(index1)).*exp(g10(index1))./(alpha1(index1).^2));

```



```

    aux2Ha2=-sum((1-alpha1(indc)).*gamma(indc).*(-c+gamma(indc).*(1-
alpha0(indc))).*exp(2*g10(indc))./(c.^2));
    aux3Ha2=sum(phi.*Hx.*(dmua1.^2)+(1-
alpha1(indc)).*exp(g10(indc)).*(yh-muh).*(-c+2*(1-
alpha1(indc)).*gamma(indc))./(c.^3));
    Ha2 = aux1Ha2+aux2Ha2+aux3Ha2;

    % second derivative a3 to a3
    aux1Ha3=-sum((1-
gamma(index1)).*exp(2*gc0(index1))./(gamma(index1).^2));
    aux2Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*phi.*dmua3.*Hx.*(1-gamma(indc))./(c.^2));
    aux3Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*(1-gamma(indc)).*(yh-
muh).*exp(gc0(indc))./(c.^2));
    aux3Ha4=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc))*2.*((1-gamma(indc)).^2).*alpha1(indc).*(yh-
muh).*exp(gc0(indc))./(c.^3));
    Ha3 = aux1Ha3+aux2Ha3+aux3Ha3+aux3Ha4;

    dna1=(1-alpha0(indc)).*(-phi.*Hx.*dmua1./(c.^2)+2*(1-
alpha1(indc)).*gamma(indc).*(yh-muh).*exp(g00(indc)))./(c.^3);
    dwa1=(1-alpha0(indc)).*(-phi.*Hx.*dmua1./(c.^2)+2*(1-
alpha0(indc)).*(1-gamma(indc)).*(yh-muh).*exp(g00(indc)))./(c.^3);
    dna2=(1-alpha0(indc)).*(-phi.*Hx.*dmua2./(c.^2)+2*(1-
alpha1(indc)).*gamma(indc).*(yh-muh).*exp(g10(indc)))./(c.^3);
    dwa2=- (1-alpha1(indc)).*(exp(g10(indc)).*(yh-
muh)+phi.*Hx.*dmua2)./(c.^2)+2*((1-alpha1(indc)).^2).*gamma(indc).*(yh-
muh).*exp(g10(indc))./(c.^3);

    % secod derivative omega-a1
    Hwa1_1= -(1-
alpha0(index0)).*exp(g00(index0))./(g01(index0).*alpha0(index0).^2);
    Hwa1_2= ((1-gamma(indc)).*(1-gamma(indc))).*(1-
alpha0(indc)).*exp(g00(indc))./((c.^2).*g01(indc));
    Hwa1_3= phi.*gamma(indc).*(1-gamma(indc)).*(1-
alpha1(indc)).*dwa1./g01(indc);
    Hwa1 = [Hwa1_1; zeros(n1,1) ;Hwa1_2+Hwa1_3];
    Hwa1 = (Xs')*(Hwa1);

    % secod derivative omega-a2
    Hwa2_1 = (1-gamma(indc)).*(1-
alpha1(indc)).*gamma(indc).*exp(g10(indc))./((c.^2).*g01(indc));
    Hwa2_2 = phi.*gamma(indc).*(1-gamma(indc)).*(dwa2)./g01(indc);
    Hwa2_ = [zeros(n0,1); zeros(n1,1) ;Hwa2_1+Hwa2_2];
    Hwa2 = (Xs')*(Hwa2_);

```

```

% secod derivative omega-a3
Hwa3_1 = -(1-gamma(indc)).*(alpha1(indc).*(1-gamma(indc))-
c).*exp(gc0(indc))./((c.^2).*g01(indc));
Hwa3_2 = phi*(1-alpha1(indc)).*(1-gamma(indc)).*(1-
2*gamma(indc)).*(yh-muh).*exp(gc0(indc))./(g01(indc));
Hwa3_3 =(phi.^2)*Hx.*dmua3.*gamma(indc).*(1-gamma(indc)).*(1-
alpha1(indc)).*exp(gc0(indc))./(g01(indc));
Hwa3_ =[zeros(n0,1); zeros(n1,1) ;Hwa3_1+Hwa3_2+Hwa3_3];
Hwa3 = (Xs')*(Hwa3_);

% secod derivative eta-a1
Hea1_1 =-(1-alpha0(indc)).*(1-
gamma(indc)).*gamma(indc).*exp(g00(indc))./((c.^2).*g11(indc));
Hea1_2 = phi.*gamma(indc).*(1-gamma(indc)).*dna1./g11(indc);
Hea1_ =[zeros(n0,1);zeros(n1,1); Hea1_1+Hea1_2];
Hea1 = (Xs')*(Hea1_);

% secod derivative eta-a2
Hea2_1 = -(1-
alpha1(index1)).*exp(g10(index1))./(alpha1(index1).*g11(index1));
Hea2_2 = -(gamma(indc).^2).*(1-
alpha1(indc)).*exp(g10(indc))./((c.^2).*g11(indc));
Hea2_3 = phi.*gamma(indc).*(1-gamma(indc)).*dna2./g11(indc);
Hea2_ =[zeros(n0,1);Hea2_1;Hea2_2+Hea2_3];
Hea2 = (Xs')*(Hea2_);

% secod derivative eta-a3
Hea3_1 = -(1-
gamma(indc)).*(c+alpha1(indc).*gamma(indc)).*exp(gc0(indc))./((c.^2).*g11(ind
c));
Hea3_2 = 2*phi*(1-alpha0(indc)).*alpha1(indc).*((1-
gamma(indc)).^2).*gamma(indc).*(yh-muh).*exp(gc0(indc))./((c.^3).*g11(indc));
Hea3_3 = -phi*(1-alpha0(indc)).*(1-gamma(indc)).*(1-
2*gamma(indc)).*(yh-muh).*exp(gc0(indc))./((c.^2).*g11(indc));
Hea3_4 = phi*phi*(1-alpha0(indc)).*(1-
gamma(indc)).*gamma(indc).*Hx.*dmua3./((c.^2).*g11(indc));
Hea3_ =[zeros(n0,1); zeros(n1,1) ;Hea3_1+Hea3_2+Hea3_3+Hea3_4];
Hea3 = (Xs')*(Hea3_);

% secod derivative beta-a1
Hba1_1 = (1-alpha1(indc)).*(c+(alpha0(indc)+alpha1(indc))).*(1-
gamma(indc)).*exp(g00(indc))./((c.^2).*gc1(indc));
Hba1_2 = phi.*(1-alpha1(indc)).*dna1./gc1(indc);
Hba1_ =[zeros(n0,1); zeros(n1,1);Hba1_1+Hba1_2];

```

```

Hba1 = (Xs')*(Hba1_);

% secod derivative beta-a2
Hba2_1 = (1-alpha1(indc)).*(-c+(alpha0(indc)-
alpha1(indc)).*gamma(indc)).*exp(g10(indc))./((c.^2).*gc1(indc));
Hba2_2 = phi.*(1-alpha0(indc)).*dna1./gc1(indc);
Hba2_ = [zeros(n0,1); zeros(n1,1);Hba2_1+Hba2_2];
Hba2 = (Xs')*(Hba2_);

% secod derivative beta-a3
Hba3_0 = exp(gc0(index0))./(1-gamma(index0)).*(gc1(index0));
Hba3_1 = -(1-
gamma(index1)).*exp(gc0(index1))./((gamma(index1).^2).*gc1(index1));
Hba3_2 = (1-gamma(indc)).*alpha1(indc).*(alpha1(indc)-
alpha0(indc)).*exp(gc0(indc))./((c.^2).*gc1(indc));
Hba3_3 = phi*(1-alpha1(indc)).*(1-alpha0(indc)).*(-
phi.*dmua3.*Hx.*(c.^2)+2*c.*alpha1(indc).*(1-gamma(indc)).*Hx.*(yh-
muh)).*exp(gc0(indc))./((c.^4).*gc1(indc));
Hba3_ = [Hba3_0; Hba3_1;Hba3_2+Hba3_3];
Hba3 = (Xs')*(Hba3_);

% secod derivative phi-a1
Hpa1_1 = gamma(indc).*(1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(g00(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa1 = sum(Hpa1_1);

% secod derivative phi-a2
Hpa2_1 = gamma(indc).*(1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(g10(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa2 = sum(Hpa2_1);

% secod derivative phi-a3
Hpa2_1 = (1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(gc0(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa3 = sum(Hpa2_1);

% second derivative a1a2
Ha1a2_1 = -sum((1-alpha0(indc)).*(1-gamma(indc)).*(1-
alpha1(indc)).*gamma(indc).*(exp(g00(indc)+g10(indc))));
Ha1a2_2 = sum(phi*(Hx.*dmua2.*dmua1+gamma(indc).*(1-
gamma(indc)).*(1-alpha0(indc)).*(1-alpha1(indc)).*c.*(gamma(indc)-
c)).*exp(2*g00(indc)));

```

```

Ha1a2 = Ha1a2_1+Ha1a2_2;

% second derivative a1a3
Ha1a3_1 = -sum((1-gamma(indc)).*(1-alpha0(indc)).*(-
c+alpha1(indc).*(1-gamma(indc))).*(exp(gc0(indc)+g10(indc))));
Ha1a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua1);
Ha1a3_3 = sum((1-alpha0(indc)).*(1-alpha1(indc)).*(((1-
gamma(indc)).^2-(1-gamma(indc)).*gamma(indc)).*(c.^2)+2*c.*(1-
gamma(indc)).*alpha1(indc)).*exp(g00(indc)+gc0(indc))));
Ha1a3 = Ha1a3_1+Ha1a3_2+Ha1a3_3;

% second derivative a2a3
Ha2a3_1 = -sum((1-
gamma(indc)).*(c+alpha1(indc).*gamma(indc)).*exp(gc0(indc))./(c.^2));
Ha2a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua2);
Ha2a3_3 = -sum((1-
gamma(indc)).*(c+2*alpha1(indc).*gamma(indc)).*(yh-muh)./(c.^3));
Ha2a3 = Ha2a3_1+Ha2a3_2+Ha2a3_3;

% Hessian Matrix
%H = -[Hphi Hphib' Hphiw0' Hphiw1';
% Hphib Hb Hbw0 Hbw1;
% Hphiw0 Hbw0 Hw0 Hbw01;
% Hphiw1 Hbw1 Hbw01 Hw1]; % Hessian matrix
% Hea1=Hea2;

H = -[Ha1 , Ha1a2 , Ha1a3 , Hpa1 , Hba1' , Hwa1' , Hea1' ;
Ha1a2 , Ha2 , Ha2a3 , Hpa2 , Hba2' , Hwa2' , Hea2' ;
Ha1a3 , Ha2a3 , Ha3 , Hpa3 , Hba3' , Hwa3' , Hea3' ;
Hpa1 , Hpa2 , Hpa3 , Hphi , Hphib' , Hphiw0' , Hphiw1';
Hba1 , Hba2 , Hba3 , Hphib , Hb , Hbw0 , Hbw1 ;
Hwa1 , Hwa2 , Hwa3 , Hphiw0 , Hbw0 , Hw0 , Hbw01 ;
Hea1 , Hea2 , Hea3 , Hphiw1 , Hbw1 , Hbw01 , Hw1 ];

end
end
end

```

9.3.6 Simulación de valores de la distribución y recuperación de parámetros caso 1 inflacionada

```
clear all;
```

```

n = 1000; % simulated data size
% Define model parameters and independent variables%%%%%%%%%
a = 0.4; %new skew parameter
phi = 100;
betag = [0.1 0.2 -0.1 0.4 0.1]';
eta = [0.05 -1 -0.5 0.5 1]';
sigmas = [0.4 -0.5 0.4 0.2; -0.5 1 0.2 0.1; 0.4 0.2 3 1.5; 0.2 0.1 1.5
1];
%sigmas=[    0.05    0.01    0.03    0.01
            %0.01    0.80    0.04    0.02
            %0.03    0.04    0.11    0.01
            %0.01    0.02    0.01    0.03];

%covarianza normal for independent variables
%sigmas = [0 0 0 0; 0 0 0 0; 0 0 0 0; 0 0 0 0];
kk = length(betag); % # of independent variables + intercept
X = mvnrnd(zeros(n, kk-1), sigmas);
X = [ones(n, 1) X];
%thetav = [a; a; a; phi ; betag ; omega ; eta]; % true parameter
thetav = [a; a; phi; betag; eta]; % true parameter
%thetav = [a; a; a; phi ; betag; eta]; % true parameter
options = optimoptions('fmincon', 'Algorithm', 'trust-region-
reflective', 'TolX', 1e-15, 'Tolfun', 1e-15, 'MaxIter', 10000, 'TolPCG', 0.0001,
'GradObj', 'on', 'Hessian', 'user-supplied', 'Display', 'off');
k = 2*kk+3; %numero de parametros
nsim = 2; % number of simulations
res = [thetav', thetav', thetav', 0, 0];
param0 = zeros(k+1, 1);
paramPW = zeros(k+1, 1);
paramBIm = zeros(k+1, 1);
warning('off', 'all');
yy = zeros(n, 1);

for j=1:nsim,
try
Y = sbeinfm_loglog_1(X, thetav);
index0 = find(Y==0);
n0 = length(index0);
index1 = find(Y==1);
n1 = length(index1);
indc = setdiff((1:n)', union(index0, index1));
Xc = X(indc, :);

```

```

Yc = Y(indc);
m = n-n0-n1;
Ym = Y+2;
% Model specification
if m==0 % Binary regression model
    model = -1;
    Ym(Ym==3)=1;
    [w, dev] =
mnrfit(X(:,2:end),Ym,'model','nominal','interactions','on');
elseif ((n0 > 0)&(n1 == 0)) % 0-inflated regression model
    model = 1;
    Ym(indc) = 1;
elseif ((n0 == 0)&(n1 > 0)) %1-inflated regression model
    model = 2;
    Ym = Ym-1;
    Ym(indc) = 1;
elseif ((n0 > 0)&(n1 > 0) & (m > 0)) %01-inflated regression model
    model = 3;
    Ym = Ym-1;
    Ym(indc) = 3;
else % beta regression model
    model = 0;
end
%----- Initial estimators-----
initheta_loglog0 %produces phi0, beta0, w0 and eta0
beta0=beta1';
%w0=w1';
eta0=eta1';
theta00 = [1,1,phi0,beta0,eta0]';
%theta00=[1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,1,-1,-1,1,1]';
%beta0=[1,1,-1,1,1];
%----- Papke-wooldridge methodology-----
tic;
[betaP,fval,exitflag] = fmincon('kfunpapke',beta0,[],[],[],[], -
1000*ones(1,kk), 1000*ones(1,kk),[], options,X,Y);
eflagP = exitflag;
%----- Two-part methodology-----
%maintwopart;
%eflag2 = exitflag;
%f2partc = -fval;

```

```

%----- Our proposal-----
mainbinf;
eflag = exitflag;
fBIm = -fval;
%t3 = toc;
res = [res; theta0', [0;0;0;betaP';zeros(kk,1)]', theta',eflag,fBIm];
yy = [yy,Y];

catch disp('Error'); %# catch the exception
continue %# Pass control to the next loop iteration
end
end

par0 = res(2:end,1:k)';
parPW = res(2:end,k+1:2*k)';
parBIm = res(2:end,2*k+1:(end-2))';
mInitial = mean(par0,2);
mPW = mean(parPW,2);
mBIm = mean(parBIm,2);
[a,ssim] = size(par0);
RSMEInitial = sqrt(mean((par0 -kron(ones(1,ssim), thetav)).^2,2));
RSMEPW = sqrt(mean((parPW -kron(ones(1,ssim), thetav)).^2,2));
RSMEBIm = sqrt(mean((parBIm -kron(ones(1,ssim), thetav)).^2,2));
%hist(Y,5);
%tabulate(Y);
table(thetav, mInitial, mPW, mBIm, RSMEInitial , RSMEPW, RSMEBIm)

```

9.3.7 Simulación de valores de la distribución y recuperación de parámetros caso 0-1 inflacionada

```

clear all;
n = 100; % simulated data size
% Define model parameters and independent variables%%%%%%%%
a = 0.5; %new skew parameter
phi = 50;
betag = [-0.5 1 -0.5 0.5 -0.5]';
eta = [0.5 -1 -0.5 0.5 1]';

```

```

    omega = [0.5 -0.5 -0.5 0.5 0.5]';
    sigmas = [0.4 -0.5 0.4 0.2; -0.5 1 0.2 0.1; 0.4 0.2 3 1.5; 0.2 0.1 1.5
1]; %covarianza normal for independent variables
    %sigmas = [0 0 0 0; 0 0 0 0; 0 0 0 0; 0 0 0 0];
    kk = length(betag); % # of independent variables + intercept
    X = mvnrnd(zeros(n, kk-1), sigmas);
    X = [ones(n, 1) X];
    thetav = [a; a; a; phi ; betag ; omega ; eta]; % true parameter
    options = optimoptions('fmincon', 'Algorithm', 'trust-region-
reflective', 'TolX', 1e-15, 'TolFun', 1e-15, 'MaxIter', 10000, 'TolPCG', 0.0001,
'GradObj', 'on', 'Hessian', 'user-supplied', 'Display', 'off');
    k = 3*kk+4;
    nsim = 2; % number of simulations
    res = [thetav', thetav', thetav', 0, 0];
    param0 = zeros(k+1, 1);
    paramPW = zeros(k+1, 1);
    paramBIm = zeros(k+1, 1);
    warning('off', 'all');
    yy = zeros(n, 1);

    for j=1:nsim,
    try
        Y = sbeinfm_loglog_01(X, thetav);
        index0 = find(Y==0);
        n0 = length(index0);
        index1 = find(Y==1);
        n1 = length(index1);
        indc = setdiff((1:n)', union(index0, index1));
        Xc = X(indc, :);
        Yc = Y(indc);
        m = n - n0 - n1;
        Ym = Y + 2;
    % Model specification
        if m==0 % Binary regression model
            model = -1;
            Ym(Ym==3)=1;
            [w, dev] =
mnrfit(X(:, 2:end), Ym, 'model', 'nominal', 'interactions', 'on');
        elseif ((n0 > 0) & (n1 == 0)) % 0-inflated regression model
            model = 1;
            Ym(indc) = 1;
        elseif ((n0 == 0) & (n1 > 0)) % 1-inflated regression model

```



```

    model = 2;
    Ym = Ym-1;
    Ym(indc) = 1;
elseif ((n0 > 0)&(n1 > 0) & (m > 0)) %01-inflated regression model
    model = 3;
    Ym = Ym-1;
    Ym(indc) = 3;
else % beta regression model
    model = 0;
end

%----- Initial estimators-----
initheta_loglog01 %produces phi0, beta0, w0 and eta0
beta0=beta1';
w0=w1';
eta0=eta1';
theta00 = [0.9,0.9,0.9,phi0,beta0,w0,eta0]';
%theta00=[0.900,0.900,0.900,105,0.678,0.798,-0.598,0.612,0.542,0.912,-
1.643,-1.443,0.523,0.534,1.123,-0.245,-0.212,0.367,0.545]';
%beta0=[0.678,0.798,-0.598,0.612,0.542];
%----- Papke-wooldridge methodology-----
tic;
[betaP,fval,exitflag] = fmincon('kfunpapke',beta0,[],[],[],[], -
1000*ones(1,kk), 1000*ones(1,kk),[], options,X,Y);
eflagP = exitflag;
%----- Two-part methodology-----
%maintwopart;
%eflag2 = exitflag;
%f2partc = -fval;
%theta00=[0.455,0.455,0.455,105.453,0.978,0.998,-
0.998,0.412,0.842,0.912,0.543,0.443,-0.523,-0.534,0.123,-0.245,-
0.212,0.367,0.545]';

%----- Our proposal-----
mainbinf;
eflag = exitflag;
fBIIm = -fval;
%t3 = toc;
res = [res; theta0', [0;0;0;0;betaP';zeros(kk+5,1)]',
theta',eflag,fBIIm];
yy = [yy,Y];

```

```

catch disp('Error'); %# Catch the exception
continue %# Pass control to the next loop iteration
end
end

par0 = res(2:end,1:k)';
parPW = res(2:end,k+1:2*k)';
parBIm = res(2:end,2*k+1:(end-2))';
mInitial = mean(par0,2);
mPW = mean(parPW,2);
mBIm = mean(parBIm,2);
[a,ssim] = size(par0);
RSMEInitial = sqrt(mean((par0 - kron(ones(1,ssim), thetav)).^2,2));
RSMEPW = sqrt(mean((parPW - kron(ones(1,ssim), thetav)).^2,2));
RSMEBIm = sqrt(mean((parBIm - kron(ones(1,ssim), thetav)).^2,2));
%hist(Y,5);
%tabulate(Y);
table(thetav, mInitial, mPW, mBIm, RSMEInitial , RSMEPW, RSMEBIm)

```

9.4 Código en Matlab CLOGLOG:

9.4.1 Función acumulada y sus derivadas

```

function [g0, g1, g2] = gloglog(y,a)
    g0 = -log(-(1./a)*log(y));
    g1 = -1./(y.*log(y));
    g2 = 1./((y.^2).*log(y)) + 1./((y.^2).*log(y).^2);

```

9.3.2 Simulación de distribuciones caso 1 inflacionado

```

if model >= 0 % initial estimation of phi and beta
    Z = log(-(1./thetav(1).*log(1-Yc)));
    %Xc= X(indc,:);
    %Ym = Y+2;
    [beta1,BINT,R,RINT,STATS] = regress(Z,Xc);
    %beta0 = inv(Xc'*Xc)*Xc'*Z;
    %ee = Z - Xc*beta0;
    %sigma2 = (ee'*ee)/(m-kk);
    %mur = 1./(1 + exp(-Xc*beta0)); logistica
    mur = 1-exp(-thetav(1)*exp(-Xc*beta1)); %caso cloglog
    mur(mur<=0.000000001) = 0.000000001;

```

```

mur(mur>=0.999999999) = 0.999999999;
sigma2 = STATS(end);
phi0x = 1./(sigma2*mur.*(1-mur));
phi0 = mean(phi0x(phi0x < 1e+4)) - 1;
end
if model ~= 0 % multinomial setting
%Ym(Ym==3)=1;
w00 = mnrfi(X(:,2:end),Ym,'model','nominal','interactions','on');
end
p0model = mnrvl(w00, X(:,2:end));
%mur2 = p0model(indc,2) + p0model(indc,3).*mur;
mur2 = (1 - p0model(indc,1)).*mur;
alpha00 = p0model(indc,1)./(1-mur2);

%alpha01 = p0model(indc,2)./mur2;
w1 = regress(log(-(1./thetav(1).*log(1-alpha00))),Xc);
%eta1 = regress(log(-(1./thetav(1).*log(1-alpha01))),Xc);

```

9.3.3 Simulación de distribuciones caso 0-1 inflacionado

```

if model >= 0 % initial estimation of phi and beta
Z = log(-(1./thetav(1).*log(1-Yc)));
%Xc= X(indc,:);
%Ym = Y+2;
[beta1,BINT,R,RINT,STATS] = regress(Z,Xc);
%beta0 = inv(Xc'*Xc)*Xc'*Z;
%ee = Z - Xc*beta0;
%sigma2 = (ee'*ee)/(m-kk);
%mur = 1./(1 + exp(-Xc*beta0)); logistica
mur = 1-exp(-thetav(1)*exp(-Xc*beta1)); %caso cloglog
mur(mur<=0.000000001) = 0.000000001;
mur(mur>=0.999999999) = 0.999999999;
sigma2 = STATS(end);
phi0x = 1./(sigma2*mur.*(1-mur));
phi0 = mean(phi0x(phi0x < 1e+4)) - 1;
end
if model ~= 0 % multinomial setting
%Ym(Ym==3)=1;
w00 = mnrfi(X(:,2:end),Ym,'model','nominal','interactions','on');
end
p0model = mnrvl(w00, X(:,2:end));
%mur2 = p0model(indc,2) + p0model(indc,3).*mur;

```

```

mur2 = p0model(indc,2) + (1 - p0model(indc,1) - p0model(indc,2)).*mur;
alpha00 = p0model(indc,1)./(1-mur2);
alpha01 = p0model(indc,2)./mur2;
w1 = regress(log(-(1./thetav(1).*log(1-alpha00))),xc);
eta1 = regress(log(-(1./thetav(1).*log(1-alpha01))),xc);

```

9.4.2 Gradiente y Matriz Hessiana para 1 inflacionado

```

function [f, grad, H] = pkfun0_cloglog(thetav,x,Y)
%global res;
[n,kk] = size(x);
phi = thetav(3);
% new parameters
a1 = thetav(1);
a3 = thetav(2);
index0 = find(Y==0);
n0 = length(index0);
indc = setdiff((1:n)',index0);
% indices modificados
pred =X*[thetav(4:kk+3) thetav(kk+4:end)];
pred = pred.*(pred<=32).*(pred>=-32) - 32*(pred < -32) + 32*(pred >
32);
F = 1-exp(-thetav(1)*exp(pred));
gamma = F(:,1);
gamma(gamma<=0.0001) = 0.0001;
gamma(gamma>=0.9999) = 0.9999;
alpha0 = F(:,2);
alpha0(alpha0<=0.0001) = 0.0001;
alpha0(alpha0>=0.9999) = 0.9999;
gc = gamma(indc);
gc(gc<=0.0001) = 0.0001;
gc(gc>=0.9999) = 0.9999;
c = 1 - alpha0(indc) + alpha0(indc).*gc;
mu = gc./c;
mu(mu<=0.0001) = 0.0001;
mu(mu>=0.9999) = 0.9999;
%mu(mu==1.0000) = 0.999;
Yc = Y(indc);
aux0 = sum(log(alpha0(index0).*(1 - gamma(index0))));
bet = betapdf(Yc,phi*mu,phi*(1-mu));
hh = c.*bet;
hh(hh <= 0.3e-323) = 0.3e-323;

```

```

aux2 = sum(log(hh));
f = -(aux0 + aux2); %- loglikelihood function
if isnan(f)==1
    f = realmax;
    grad = ones(2*kk + 1,1);
    H = ones(2*kk+1,2*kk+1);
else
    if nargout > 1 % with gradient

        [g00, g01, g02] = cgloglog(alpha0, a1);
        [gc0, gc1, gc2] = cgloglog(gamma, a3);
        % dim(Xs) = dim(X)

        Xs = [X(index0,:) ; X(indc,:)];

        g00s = [g00(index0); g00(indc)];
        gc0s = [gc0(index0); gc0(indc)];
        g01s = [g01(index0); g01(indc)];
        gc1s = [gc1(index0); gc1(indc)];
        yh = log(Yc./(1-Yc));
        muh = psi(phi*mu) - psi(phi*(1-mu));
        gradphi = sum(mu.*(yh - muh) + log(1-Yc) - psi(phi*(1-mu)) +
psi(phi));
        Mg = [ 1./(gamma(index0)-1) ; alpha0(indc)./c + phi*(1-
alpha0(indc)).*(yh-muh)./(c.^2)];
        gradg = (Xs')*(Mg./gc1s);
        Mg0 = [1./alpha0(index0) ; (gc-1)./c + (phi*gc.*(1-gc).*(yh-
muh))./(c.^2)];
        grad0 = (Xs')*(Mg0./g01s);
        dmua1=gc.*(1-gc).*(1-alpha0(indc)).*exp(g00(indc))./(c.^2);
        dmua3=(1-alpha0(indc)).*(1-gc).*exp(gc0(indc))./(c.^2);
        dmua11=gc.*(1-gc).*(1-alpha0(indc))./(c.^2);%exp(g00(indc))
        dmua31=(1-alpha0(indc)).*(1-gc)./(c.^2);%exp(gc0(indc))
        %grada1 = sum((1-alpha0(index0)).*g00(index0)./alpha0(index0))-
sum((1-alpha0(indc)).*(1-gc).*g00(indc)./c) +sum(dmua1.*phi.*(yh - muh));
        grada1 = [(1-alpha0(index0))./alpha0(index0);-(1-alpha0(indc)).*(1-
gc)./c + dmua11.*phi.*(yh - muh)];
        grada11=sum(grada1.*exp(g00s));
        %grada3 = sum(-gc0(index0))+ sum((1-
gamma(index1)).*gc0(index1))./gamma(index1) + sum(dmua3.*phi.*(yh - muh));
        grada3 = [-ones(n0,1);dmua31.*phi.*(yh - muh)];
        grada33=sum(grada3.*exp(gc0s));

```

```

grad = -[grada11;grada33;gradphi ; gradg ; grad0]; % gradient
if nargout > 2 % with Hessian
    Hx = psi(1, phi*mu) + psi(1,phi*(1-mu));
    Hphi = (n - n0)*psi(1,phi) - sum((mu.^2).*psi(1,phi*mu)+ ((1 -
mu).^2).*psi(1,phi*(1-mu)));
    auxx1 = yh-muh-phi*mu.*psi(1,phi*mu)+phi*(1-mu).*psi(1,(1-mu)*phi);
    %second derivate w.r.t to beta and phi
    auxphib = (1-alpha0(indc)).*auxx1./(gc1(indc).*(c.^2));
    Hphib = (X(indc,:))'*auxphib;
    %second derivate w.r.t to beta
    auxb1 = -alpha0(indc).*(alpha0(indc).*gc1(indc) +
c.*gc2(indc))./((c.^2).*(gc1(indc).^3));
    auxb2 = - (phi^2)*Hx.*((1 -
alpha0(indc)).^2)./((c.^4).*(gc1(indc).^2));
    auxb3 = phi*(alpha0(indc)-1).*(yh-muh).*(c.*gc2(indc) +
2*alpha0(indc).*gc1(indc))./((c.^3).*(gc1(indc).^3));
    auxb = [(gc2(index0).*(1-gamma(index0))- gc1(index0))./(((1 -
gamma(index0)).^2).*(gc1(index0).^3)) ; auxb1 + auxb2 + auxb3];
    auxb = sparse(auxb);
    B = diag(auxb);
    Hb = Xs'*B*Xs;
    %second derivate w.r.t to omega and phi
    auxphiw0 = gc.*(1-gc).*auxx1./(g01(indc).*(c.^2));
    Hphiw0 = (X(indc,:))'*auxphiw0;
    %second derivate w.r.t to beta and omega
    auxx2 = 1 - alpha0(indc) + gc.*(alpha0(indc) -2);
    auxbw01 = 1./((c.^2).*g01(indc).*gc1(indc));
    auxbw02 = phi*(yh - muh).*auxx2./((c.^3).*g01(indc).*gc1(indc));
    auxbw03 = -(phi^2).*((1 - alpha0(indc)).^2).*gc.*(1-
gc).*Hx./((c.^4).*g01(indc).*gc1(indc));
    auxbw0 = auxbw01 + auxbw02 + auxbw03;
    auxbw0 = sparse(auxbw0);
    Bw0 = diag(auxbw0);
    Hbw0 = X(indc,:))'*Bw0*X(indc,:);
    %second derivate w.r.t to omega
    auxw01 = ((gc-1).*g01(indc) + c.*g02(indc)).*(1-
gc)./((c.^2).*(g01(indc).^3));
    auxw02 = -(phi^2).*(gc.^2).*((1-
gc).^2).*Hx./((c.^4).*(g01(indc).^2));
    auxw03 = -phi*gc.*(1 -gc).*(yh -muh).*(c.*g02(indc) - 2*(1-
gc).*g01(indc))./((c.^3).*(g01(indc).^3));
    auxw0 = [-(g01(index0) +
alpha0(index0).*g02(index0))./((alpha0(index0).^2).*(g01(index0).^3)); auxw01
+ auxw02+ auxw03];
    auxw0 = sparse(auxw0);

```

```

    Bw0 = diag(auxw0);
    Hw0 = Xs'*Bw0*Xs;
    % second derivative a1 to a1
    aux1Ha1=-sum((1-
alpha0(index0)).*exp(2*g00(index0))./(alpha0(index0).^2));
    aux2Ha1=sum((1-alpha0(indc)).*(1-gamma(indc)).*(c+(1-
gamma(indc)).*alpha0(indc)).*exp(2*g00(indc))./(c.^2));
    aux3Ha1=-sum(phi.*Hx.*(dmua1.^2)+(1-
alpha0(indc)).*exp(g00(indc)).*(yh-muh).*(c-2*(1-alpha0(indc)).*(1-
gamma(indc)))./(c.^3));
    Ha1 = aux1Ha1+aux2Ha1+aux3Ha1;
    % second derivative a3 to a3
    aux2Ha3=-sum(phi.*exp(gc0(indc)).*phi.*dmua3.*Hx.*(1-
gamma(indc))./(c.^2));
    aux3Ha3=-sum(phi.*exp(gc0(indc)).*(1-gamma(indc)).*(yh-
muh).*exp(gc0(indc))./(c.^2));
    aux3Ha4=-sum(phi.*exp(gc0(indc))*2.*((1-gamma(indc)).^2).*(yh-
muh).*exp(gc0(indc))./(c.^3));
    Ha3 = aux2Ha3+aux3Ha3+aux3Ha4;
    % second derivative a1a3
    Ha1a3_1 = -sum((1-gamma(indc)).*(1-alpha0(indc)).*(1-
gamma(indc)).*(exp(gc0(indc)))));
    Ha1a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua1);
    Ha1a3_3 = sum((1-alpha0(indc)).*(((1-gamma(indc)).^2-(1-
gamma(indc)).*gamma(indc)).*(c.^2)+2*c.*(1-
gamma(indc)).*exp(g00(indc)+gc0(indc))));
    Ha1a3 = Ha1a3_1+Ha1a3_2+Ha1a3_3;
    % secod derivative phi-a1
    Hpa1_1 = gamma(indc).*(1-gamma(indc)).*(1-
alpha0(indc)).*(exp(g00(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
    Hpa1 = sum(Hpa1_1);
    % secod derivative phi-a3
    Hpa2_1 = (1-gamma(indc)).*(1-alpha0(indc)).*(exp(gc0(indc))).*((yh-
muh)-phi*Hx.*mu+phi*psi(1,phi*(1-mu)))./(c.^2);
    Hpa3 = sum(Hpa2_1);
    % secod derivative omega-a3
    Hwa3_1 = -(1-gamma(indc)).*((1-gamma(indc))-
c).*exp(gc0(indc))./(c.^2).*g01(indc));
    Hwa3_2 = phi*(1-gamma(indc)).*(1-2*gamma(indc)).*(yh-
muh).*exp(gc0(indc))./(g01(indc));
    Hwa3_3 =(phi.^2)*Hx.*dmua3.*gamma(indc).*(1-
gamma(indc)).*exp(gc0(indc))./(g01(indc));
    Hwa3_ = [zeros(n0,1);Hwa3_1+Hwa3_2+Hwa3_3];
    Hwa3 = (Xs')*(Hwa3_);
    dna1=(1-alpha0(indc)).*(-phi*Hx.*dmua1./(c.^2)+2*gamma(indc).*(yh-
muh).*exp(g00(indc)))./(c.^3);

```

```

    dwa1=(1-alpha0(indc)).*(-phi*Hx.*dmua1./(c.^2)+2*(1-alpha0(indc)).*(1-
gamma(indc)).*(yh-muh).*exp(g00(indc)))./(c.^3);
    % secod derivative omega-a1
    Hwa1_1= -(1-
alpha0(index0)).*exp(g00(index0))./(g01(index0).*alpha0(index0).^2);
    Hwa1_2= ((1-gamma(indc)).*(1-gamma(indc))).*(1-
alpha0(indc)).*exp(g00(indc))./((c.^2).*g01(indc));
    Hwa1_3= phi*gamma(indc).*(1-gamma(indc)).*dwa1./(g01(indc));
    Hwa1 = [Hwa1_1;Hwa1_2+Hwa1_3];
    Hwa1 = (Xs')*(Hwa1);
    % secod derivative beta-a1
    Hba1_1 = (c+alpha0(indc)).*(1-
gamma(indc)).*exp(g00(indc))./((c.^2).*gc1(indc));
    Hba1_2 = phi.*dna1./gc1(indc);
    Hba1_ = [zeros(n0,1);Hba1_1+Hba1_2];
    Hba1 = (Xs')*(Hba1_);
    % secod derivative beta-a3
    Hba3_0 = exp(gc0(index0))./(1-gamma(index0)).*(gc1(index0));
    Hba3_2 = (1-
gamma(indc)).*alpha0(indc).*exp(gc0(indc))./((c.^2).*gc1(indc));
    Hba3_3 = phi*(1-alpha0(indc)).*(-phi.*dmua3.*Hx.*(c.^2)+2*c.*(1-
gamma(indc)).*Hx.*(yh-muh).*exp(gc0(indc)))./((c.^4).*gc1(indc));
    Hba3_ = [Hba3_0;Hba3_2+Hba3_3];
    Hba3 = (Xs')*(Hba3_);

    %H = -[ Hphi Hphib' Hphiw0' ;
           %Hphib Hb Hbw0 ;
           %Hphiw0 Hbw0 Hw0];
           % Hessian matrix

    H = -[Ha1    , Ha1a3 , Hpa1    , Hba1'  , Hwa1'  ;
          Ha1a3 , Ha3    , Hpa3    , Hba3'  , Hwa3'  ;
          Hpa1  , Hpa3   , Hphi    , Hphib' , Hphiw0' ;
          Hba1  , Hba3   , Hphib   , Hb     , Hbw0   ;
          Hwa1  , Hwa3   , Hphiw0  , Hbw0  , Hw0    ];

    end
end
end
%res = vertcat(res,[theta',f]);

```

9.4.3 Gradiente y Hesiiana para 0-1 Inflacionado

```
function [f, grad, H] = pkfun01_cloglog(thetav,X,Y)
```



```

[n, kk] = size(X);
phi = thetav(4);
a1 = thetav(1);
a2 = thetav(2);
a3 = thetav(3);
index0 = find(Y == 0.0000);
n0 = length(index0);
index1 = find(Y == 1.0000);
n1 = length(index1);
indc = union(index0,index1);
indc = setdiff((1:n)',indc);
pred = X*[thetav(5:kk+4) thetav(kk+5:2*kk+4) thetav(2*kk+5:end)];
pred = pred.*(pred<=12).*(pred>=-700) - 700*(pred < -700) +
12*(pred > 12);
% function loglog
%F = exp(-thetav(1)*exp(-pred));
% function cloglog
F = 1-exp(-thetav(1)*exp(pred));
gamma = F(:,1);
gamma(gamma<=0.0001) = 0.0001;
gamma(gamma>=0.9999) = 0.9999;
alpha0 = F(:,2);
alpha0(alpha0<=0.0001) = 0.0001;
alpha0(alpha0>=0.9999) = 0.9999;
alpha1 = F(:,3);
alpha1(alpha1<=0.0001) = 0.0001;
alpha1(alpha1>=0.9999) = 0.9999;
gc = gamma(indc);
gc(gc<=0.0001) = 0.0001;
gc(gc>=0.9999) = 0.9999;
c = 1 - alpha0(indc) + (alpha0(indc) - alpha1(indc)).*gc;
mu = gc.*(1- alpha1(indc))./c;
mu(mu<=0.0001) = 0.0001;
mu(mu>=0.9999) = 0.9999;
Yc = Y(indc);
aux0 = sum(log(alpha0(index0).*(1 - gamma(index0))));
aux1 = sum(log(alpha1(index1).*gamma(index1)));
bet = betapdf(Yc,phi*mu,phi*(1-mu));
hh = c.*bet;
hh(hh <= 0.3e-323) = 0.3e-323;
aux2 = sum(log(hh));

```

```

f = -(aux0 + aux1+ aux2); %- loglikelihood function
if isnan(f)==1
    f = realmax;
    grad = ones(3*kk+4,1);
    H = ones(3*kk+4,3*kk+4);
else

if nargout > 1 % with gradient
    [g00, g01, g02] = cgloglog(alpha0, a1);
    [g10, g11, g12] = cgloglog(alpha1, a2);
    [gc0, gc1, gc2] = cgloglog(gamma, a3);
    % dim(Xs) = dim(X)
    Xs = [X(index0,:) ; X(index1,:) ; X(indc,:)];
    g00s = [g00(index0); g00(index1); g00(indc)];
    g10s = [g10(index0); g10(index1); g10(indc)];
    gc0s = [gc0(index0); gc0(index1); gc0(indc)];
    g01s = [g01(index0); g01(index1); g01(indc)];
    g11s = [g11(index0); g11(index1); g11(indc)];
    gc1s = [gc1(index0); gc1(index1); gc1(indc)];
    yh = log(Yc./(1-Yc));
    muh = psi(phi*mu) - psi(phi*(1-mu));
    gradphi = sum(mu.*(yh - muh) + log(1-Yc) - psi(phi*(1-mu)) +
psi(phi));
    Mg = [ 1./(gamma(index0)-1) ; 1./gamma(index1) ; (alpha0(indc)-
alpha1(indc))./c + phi*(1-alpha0(indc)).*(1-alpha1(indc)).*(yh-muh)./(c.^2)];
    gradg = (Xs')*(Mg./gc1s);
    Mg0 = [1./alpha0(index0) ; zeros(n1,1) ; (gc-1)./c +
(phi*gc.*(1-gc)).*(1-alpha1(indc)).*(yh-muh))./(c.^2)];
    grad0 = (Xs')*(Mg0./g01s);
    Mg1 = [zeros(n0,1) ; 1./alpha1(index1) ; -gc./c + (phi*gc.*(1-
gc).*(alpha0(indc)-1).*(yh-muh))./(c.^2)];
    grad1 = (Xs')*(Mg1./g11s);
    dmua1=gc.*(1-gc).*(1-alpha1(indc)).*(1-
alpha0(indc)).*exp(g00(indc))./(c.^2);
    dmua2=-gc.*(1-gc).*(1-alpha0(indc)).*(1-
alpha1(indc)).*exp(g10(indc))./(c.^2);
    dmua3=(1-alpha1(indc)).*(1-alpha0(indc)).*(1-
gc).*exp(gc0(indc))./(c.^2);
    dmua11=gc.*(1-gc).*(1-alpha1(indc)).*(1-
alpha0(indc))./(c.^2);%exp(g00(indc))
    dmua21=-gc.*(1-gc).*(1-alpha0(indc)).*(1-
alpha1(indc))./(c.^2);%exp(g10(indc))
    dmua31=(1-alpha1(indc)).*(1-alpha0(indc)).*(1-
gc)./(c.^2);%exp(gc0(indc))

```

```

        %grada1 = sum((1-alpha0(index0)).*g00(index0)./alpha0(index0))-
sum((1-alpha0(indc)).*(1-gc).*g00(indc)./c) +sum(dmua1.*phi.*(yh - muh));
        grada1 = [(1-alpha0(index0))./alpha0(index0);zeros(n1,1);-(1-
alpha0(indc)).*(1-gc)./c + dmua11.*phi.*(yh - muh)];
        grada11=sum(grada1.*exp(g00s));
        %grada2 = sum((1-alpha1(index1)).*g01(index1)./alpha1(index1))-
sum((1-alpha1(indc)).*gc.*g01(indc)./c) + sum(dmua2.*phi.*(yh - muh));
        grada2 = [zeros(n0,1);(1-alpha1(index1))./alpha1(index1);-(1-
alpha1(indc)).*gc./c-dmua21.*phi.*(yh - muh)];
        grada22=sum(grada2.*exp(g10s));
        %grada3 = sum(-gc0(index0))+ sum((1-
gamma(index1)).*gc0(index1))./gamma(index1) + sum(dmua3.*phi.*(yh - muh));
        grada3 = [-ones(n0,1);(1-
gamma(index1))./gamma(index1);dmua31.*phi.*(yh - muh)];
        grada33=sum(grada3.*exp(gc0s));
        grad = -[grada11; grada22; grada33; gradphi; gradg ; grad0 ;
grad1]; % gradiente
    end
    if nargout > 2 % with Hessian
        Hx = psi(1, phi*mu) + psi(1,phi*(1-mu));
        Hphi = (n - n0 - n1)*psi(1,phi) - sum((mu.^2).*psi(1,phi*mu)+
((1 - mu).^2).*psi(1,phi*(1-mu)));
        auxx1 = yh-muh-phi*mu.*psi(1,phi*mu)+phi*(1-mu).*psi(1,(1-
mu)*phi);
        %second derivate w.r.t to beta and phi
        auxphib = (1-alpha0(indc)).*(1-
alpha1(indc)).*auxx1./(gc1(indc).*(c.^2));
        Hphib = (X(indc,:))'*auxphib;
        %second derivate w.r.t to beta
        %cg = c.*gc1(indc);
        %cg = cg.*(cg<=5.6438e+102) + 5.6438e+102.*(cg > 5.6438e+102);
        auxb1 = -(alpha0(indc)-alpha1(indc)).*((alpha0(indc)-
alpha1(indc)).*gc1(indc) + c.*gc2(indc))./((c.^2).*(gc1(indc).^3));
        auxb2 = - ((phi^2)*Hx.*((1 -alpha0(indc)).^2).*((1-
alpha1(indc)).^2))./((c.^4).*(gc1(indc).^2));
        nauxb3 = -phi*(1-alpha0(indc)).*(1- alpha1(indc)).*(yh-
muh).*(c.*gc2(indc) + 2*(alpha0(indc)- alpha1(indc)).*gc1(indc));
        dauxb3 = (c.^3).*(gc1(indc).^3);
        nauxb3(nauxb3 ==Inf) = realmax;
        dauxb3(dauxb3 ==Inf) = realmax;
        auxb3 = nauxb3./dauxb3;
        auxb = [(gc2(index0).*(1-gamma(index0))- gc1(index0))./(((1 -
gamma(index0)).^2).*(gc1(index0).^3)) ; -(gc2(index1).*gamma(index1) +
gc1(index1))./((gamma(index1).^2).*(gc1(index1).^3)) ; auxb1 + auxb2 +
auxb3];
        %auxb(auxb==Inf|isnan(auxb)==1) = 0

```

```

        %[-phi*(1-alpha0(indc)).*(1- alpha1(indc)).*(yh-
muh).*(c.*gc2(indc) + 2*(alpha0(indc)- alpha1(indc)).*gc1(indc))
(c.^3).*(gc1(indc).^3) auxb3]
        auxb = sparse(auxb);
        B = diag(auxb);
        Hb = Xs'*B*Xs;
        %Hb(Hb==Inf|isnan(Hb)==1) = 0
        %second derivate w.r.t to omega and phi
        auxphiw0 = gc.*(1-gc).*(1-
alpha1(indc)).*auxx1./(g01(indc).*(c.^2));
        Hphiw0 = (X(indc,:))'*auxphiw0;
        %second derivate w.r.t to eta and phi
        auxphiw1 = gc.*(1-gc).*(alpha0(indc)-
1).*auxx1./(g11(indc).*(c.^2));
        Hphiw1 = (X(indc,:))'*auxphiw1;
        %second derivate w.r.t to beta and omega
        auxx2 = 1 - alpha0(indc) + gc.*(alpha0(indc) + alpha1(indc) -2);
        auxbw01 = (1-alpha1(indc))./((c.^2).*g01(indc).*gc1(indc));
        auxbw02 = phi*(1- alpha1(indc)).*(yh -
muh).*auxx2./((c.^3).*g01(indc).*gc1(indc));
        auxbw03 = -(phi^2).*(1 - alpha0(indc)).*((1-
alpha1(indc)).^2).*gc.*(1- gc).*Hx./((c.^4).*g01(indc).*gc1(indc));
        auxbw0 = auxbw01 + auxbw02 + auxbw03;
        auxbw0 = sparse(auxbw0);
        Bw0 = diag(auxbw0);
        Hbw0 = X(indc,:)'*Bw0*X(indc,:);

        %second derivate w.r.t to beta and eta
        auxbw11 = (alpha0(indc)-1)./((c.^2).*g11(indc).*gc1(indc));
        auxbw12 = -phi*(1- alpha0(indc)).*(yh-
muh).*auxx2./((c.^3).*g11(indc).*gc1(indc));
        auxbw13 = (phi^2).*((1 - alpha0(indc)).^2).*(1 -
alpha1(indc)).*gc.*(1- gc).*Hx./((c.^4).*g11(indc).*gc1(indc));
        auxbw1 = auxbw11 + auxbw12 + auxbw13;
        auxbw1 = sparse(auxbw1);
        Bw1 = diag(auxbw1);
        Hbw1 = X(indc,:)'*Bw1*X(indc,:);

        %second derivate w.r.t to omega
        auxw01 = ((gc-1).*g01(indc) + c.*g02(indc)).*(1-
gc)./((c.^2).*(g01(indc).^3));
        auxw02 = -(phi^2).*(gc.^2).*((1-gc).^2).*((1-
alpha1(indc)).^2).*Hx./((c.^4).*(g01(indc).^2));
        auxw03 = -phi*gc.*(1 -gc).*(1 - alpha1(indc)).*(yh -
muh).*(c.*g02(indc) - 2*(1-gc).*g01(indc))./((c.^3).*(g01(indc).^3));

```

```

    auxw0 = [-(g01(index0) +
alpha0(index0).*g02(index0))./((alpha0(index0).^2).*(g01(index0).^3));
zeros(n1,1); auxw01 + auxw02+ auxw03];
    auxw0 = sparse(auxw0);
    Bw0 = diag(auxw0);
    Hw0 = Xs'*Bw0*Xs;
    %second derivate w.r.t to eta
    auxw11 = (c.*g12(indc)- gc.*g11(indc)
).*gc./((c.^2).*(g11(indc).^3));
    auxw12 = -(phi^2)*(gc.^2).*((1-gc).^2).*((1-
alpha0(indc)).^2).*Hx./((c.^4).*(g11(indc).^2));
    auxw13 = phi*gc.*(1-gc).*(1- alpha0(indc)).*(yh-
muh).*(c.*g12(indc)- 2*gc.*g11(indc))./((c.^3).*(g11(indc).^3));
    auxw1 = [zeros(n0,1); -(g11(index1) +
alpha1(index1).*g12(index1))./((alpha1(index1).^2).*(g11(index1).^3)); auxw11
+ auxw12 + auxw13];
    auxw1 = sparse(auxw1);
    Bw1 = diag(auxw1);
    Hw1 = Xs'*Bw1*Xs;
    %second derivate w.r.t to eta and omega
    auxw011 = gc.*(gc-1)./((c.^2).*g01(indc).*g11(indc));
    auxw012 = -phi*gc.*(1-gc).*(yh-
muh).*auxx2./((c.^3).*g01(indc).*g11(indc));
    auxw013 = (phi^2)*(1- alpha0(indc)).*(1-
alpha1(indc)).*(gc.^2).*((1-gc).^2).*Hx./((c.^4).*g01(indc).*g11(indc));
    auxw01 = auxw011 + auxw012 + auxw013;
    auxw01 = sparse(auxw01);
    Bw01 = diag(auxw01);
    Hbw01 = X(indc,:)'*Bw01*X(indc,:);
    % second derivative a1 to a1
    aux1Ha1=-sum((1-
alpha0(index0)).*exp(2*g00(index0))./(alpha0(index0).^2));
    aux2Ha1=sum((1-alpha0(indc)).*(1-gamma(indc)).*(c+(1-
gamma(indc)).*alpha0(indc)).*exp(2*g00(indc))./(c.^2));
    aux3Ha1=-sum(phi.*Hx.*(dmua1.^2)+(1-
alpha0(indc)).*exp(g00(indc)).*(yh-muh).*(c-2*(1-alpha0(indc)).*(1-
gamma(indc)))./(c.^3));
    Ha1 = aux1Ha1+aux2Ha1+aux3Ha1;
    % second derivative a2 to a2
    aux1Ha2=-sum((1-
alpha0(index1)).*exp(g10(index1))./(alpha1(index1).^2));
    aux2Ha2=-sum((1-alpha1(indc)).*gamma(indc).*(-c+gamma(indc).*(1-
alpha0(indc))).*exp(2*g10(indc))./(c.^2));
    aux3Ha2=sum(phi.*Hx.*(dmua1.^2)+(1-
alpha1(indc)).*exp(g10(indc)).*(yh-muh).*(-c+2*(1-
alpha1(indc)).*gamma(indc))./(c.^3));
    Ha2 = aux1Ha2+aux2Ha2+aux3Ha2;

```

```

% second derivative a3 to a3
    aux1Ha3=-sum((1-
gamma(index1)).*exp(2*gc0(index1))./(gamma(index1).^2));
    aux2Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*phi.*dmua3.*Hx.*(1-gamma(indc))./(c.^2));
    aux3Ha3=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc)).*(1-gamma(indc)).*(yh-
muh).*exp(gc0(indc))./(c.^2));
    aux3Ha4=-sum(phi.*(1-alpha1(indc)).*(1-
alpha1(indc)).*exp(gc0(indc))*2.*((1-gamma(indc)).^2).*alpha1(indc).*(yh-
muh).*exp(gc0(indc))./(c.^3));
    Ha3 = aux1Ha3+aux2Ha3+aux3Ha3+aux3Ha4;
    dna1=(1-alpha0(indc)).*(-phi.*Hx.*dmua1./(c.^2)+2*(1-
alpha1(indc)).*gamma(indc).*(yh-muh).*exp(g00(indc)))./(c.^3);
    dwa1=(1-alpha0(indc)).*(-phi.*Hx.*dmua1./(c.^2)+2*(1-
alpha0(indc)).*(1-gamma(indc)).*(yh-muh).*exp(g00(indc)))./(c.^3);
    dna2=(1-alpha0(indc)).*(-phi.*Hx.*dmua2./(c.^2)+2*(1-
alpha1(indc)).*gamma(indc).*(yh-muh).*exp(g10(indc)))./(c.^3);
    dwa2=- (1-alpha1(indc)).*(exp(g10(indc)).*(yh-
muh)+phi.*Hx.*dmua2)./(c.^2)+2*((1-alpha1(indc)).^2).*gamma(indc).*(yh-
muh).*exp(g10(indc))./(c.^3);
    % secod derivative omega-a1
    Hwa1_1= -(1-
alpha0(index0)).*exp(g00(index0))./(g01(index0).*alpha0(index0).^2);
    Hwa1_2= ((1-gamma(indc)).*(1-gamma(indc))).*(1-
alpha0(indc)).*exp(g00(indc))./((c.^2).*g01(indc));
    Hwa1_3= phi.*gamma(indc).*(1-gamma(indc)).*(1-
alpha1(indc)).*dwa1./(g01(indc));
    Hwa1 = [Hwa1_1; zeros(n1,1) ;Hwa1_2+Hwa1_3];
    Hwa1 = (Xs')*(Hwa1);
    % secod derivative omega-a2
    Hwa2_1 = (1-gamma(indc)).*(1-
alpha1(indc)).*gamma(indc).*exp(g10(indc))./((c.^2).*g01(indc));
    Hwa2_2 = phi.*gamma(indc).*(1-gamma(indc)).*(dwa2)./(g01(indc));
    Hwa2_ = [zeros(n0,1); zeros(n1,1) ;Hwa2_1+Hwa2_2];
    Hwa2 = (Xs')*(Hwa2_);
    % secod derivative omega-a3
    Hwa3_1 = -(1-gamma(indc)).*(alpha1(indc).*(1-gamma(indc))-
c).*exp(gc0(indc))./((c.^2).*g01(indc));
    Hwa3_2 = phi.*(1-alpha1(indc)).*(1-gamma(indc)).*(1-
2*gamma(indc)).*(yh-muh).*exp(gc0(indc))./(g01(indc));
    Hwa3_3 =(phi.^2)*Hx.*dmua3.*gamma(indc).*(1-gamma(indc)).*(1-
alpha1(indc)).*exp(gc0(indc))./(g01(indc));
    Hwa3_ = [zeros(n0,1); zeros(n1,1) ;Hwa3_1+Hwa3_2+Hwa3_3];
    Hwa3 = (Xs')*(Hwa3_);
    % secod derivative eta-a1
    Hea1_1 =-(1-alpha0(indc)).*(1-
gamma(indc)).*gamma(indc).*exp(g00(indc))./((c.^2).*g11(indc));

```

```

Hea1_2 = phi.*gamma(indc).*(1-gamma(indc)).*dna1./g11(indc);
Hea1_ = [zeros(n0,1);zeros(n1,1); Hea1_1+Hea1_2];
Hea1 = (Xs')*(Hea1_);

% secod derivative eta-a2
Hea2_1 = -(1-
alpha1(index1)).*exp(g10(index1))./(alpha1(index1).*g11(index1));
Hea2_2 = -(gamma(indc).^2).*(1-
alpha1(indc)).*exp(g10(indc))./((c.^2).*g11(indc));
Hea2_3 = phi.*gamma(indc).*(1-gamma(indc)).*dna2./g11(indc);
Hea2_ = [zeros(n0,1);Hea2_1;Hea2_2+Hea2_3];
Hea2 = (Xs')*(Hea2_);
% secod derivative eta-a3
Hea3_1 = -(1-
gamma(indc)).*(c+alpha1(indc).*gamma(indc)).*exp(gc0(indc))./((c.^2).*g11(ind
c));
Hea3_2 = 2*phi*(1-alpha0(indc)).*alpha1(indc).*((1-
gamma(indc)).^2).*gamma(indc).*(yh-muh).*exp(gc0(indc))./((c.^3).*g11(indc));
Hea3_3 = -phi*(1-alpha0(indc)).*(1-gamma(indc)).*(1-
2*gamma(indc)).*(yh-muh).*exp(gc0(indc))./((c.^2).*g11(indc));
Hea3_4 = phi*phi*(1-alpha0(indc)).*(1-
gamma(indc)).*gamma(indc).*Hx.*dmua3./((c.^2).*g11(indc));
Hea3_ = [zeros(n0,1); zeros(n1,1) ;Hea3_1+Hea3_2+Hea3_3+Hea3_4];
Hea3 = (Xs')*(Hea3_);
% secod derivative beta-a1
Hba1_1 = (1-alpha1(indc)).*(c+(alpha0(indc)+alpha1(indc))).*(1-
gamma(indc)).*exp(g00(indc))./((c.^2).*gc1(indc));
Hba1_2 = phi.*(1-alpha1(indc)).*dna1./gc1(indc);
Hba1_ = [zeros(n0,1); zeros(n1,1);Hba1_1+Hba1_2];
Hba1 = (Xs')*(Hba1_);
% secod derivative beta-a2
Hba2_1 = (1-alpha1(indc)).*(-c+(alpha0(indc)-
alpha1(indc)).*gamma(indc)).*exp(g10(indc))./((c.^2).*gc1(indc));
Hba2_2 = phi.*(1-alpha0(indc)).*dna1./gc1(indc);
Hba2_ = [zeros(n0,1); zeros(n1,1);Hba2_1+Hba2_2];
Hba2 = (Xs')*(Hba2_);
% secod derivative beta-a3
Hba3_0 = exp(gc0(index0))./(1-gamma(index0)).*(gc1(index0));
Hba3_1 = -(1-
gamma(index1)).*exp(gc0(index1))./((gamma(index1).^2).*gc1(index1));
Hba3_2 = (1-gamma(indc)).*alpha1(indc).*(alpha1(indc)-
alpha0(indc)).*exp(gc0(indc))./((c.^2).*gc1(indc));
Hba3_3 = phi*(1-alpha1(indc)).*(1-alpha0(indc)).*(-
phi.*dmua3.*Hx.*(c.^2)+2*c.*alpha1(indc).*(1-gamma(indc)).*Hx.*(yh-
muh).*exp(gc0(indc)))./((c.^4).*gc1(indc));

```

```

Hba3_ = [Hba3_0; Hba3_1; Hba3_2+Hba3_3];
Hba3 = (Xs')*(Hba3_);
% second derivative phi-a1
Hpa1_1 = gamma(indc).*(1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(g00(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa1 = sum(Hpa1_1);
% second derivative phi-a2
Hpa2_1 = gamma(indc).*(1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(g10(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa2 = sum(Hpa2_1);
% second derivative phi-a3
Hpa2_1 = (1-gamma(indc)).*(1-alpha1(indc)).*(1-
alpha0(indc)).*(exp(gc0(indc))).*((yh-muh)-phi*Hx.*mu+phi*psi(1,phi*(1-
mu)))./(c.^2);
Hpa3 = sum(Hpa2_1);

% second derivative a1a2
Ha1a2_1 = -sum((1-alpha0(indc)).*(1-gamma(indc)).*(1-
alpha1(indc)).*gamma(indc).*(exp(g00(indc)+g10(indc))));
Ha1a2_2 = sum(phi*(Hx.*dmua2.*dmua1+gamma(indc).*(1-
gamma(indc)).*(1-alpha0(indc)).*(1-alpha1(indc)).*c.*(gamma(indc)-
c)).*exp(2*g00(indc)));
Ha1a2 = Ha1a2_1+Ha1a2_2;
% second derivative a1a3
Ha1a3_1 = -sum((1-gamma(indc)).*(1-alpha0(indc)).*(-
c+alpha1(indc).*(1-gamma(indc))).*(exp(gc0(indc)+g10(indc))));
Ha1a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua1);
Ha1a3_3 = sum((1-alpha0(indc)).*(1-alpha1(indc)).*(((1-
gamma(indc)).^2-(1-gamma(indc)).*gamma(indc)).*(c.^2)+2*c.*(1-
gamma(indc)).*alpha1(indc)).*exp(g00(indc)+gc0(indc)));
Ha1a3 = Ha1a3_1+Ha1a3_2+Ha1a3_3;
% second derivative a2a3
Ha2a3_1 = -sum((1-
gamma(indc)).*(c+alpha1(indc).*gamma(indc)).*exp(gc0(indc))./(c.^2));
Ha2a3_2 = -sum(phi*phi*Hx.*dmua3.*dmua2);
Ha2a3_3 = -sum((1-
gamma(indc)).*(c+2*alpha1(indc).*gamma(indc)).*(yh-muh)./(c.^3));
Ha2a3 = Ha2a3_1+Ha2a3_2+Ha2a3_3;
% Hessian Matrix
%H = -[Hphi Hphib' Hphiw0' Hphiw1';
% Hphib Hb Hbw0 Hbw1;
% Hphiw0 Hbw0 Hw0 Hbw01;
% Hphiw1 Hbw1 Hbw01 Hw1]; % Hessian matrix
% Hea1=Hea2;

```



```

H = -[Ha1    , Ha1a2 , Ha1a3 , Hpa1    , Hba1' , Hwa1' , Hea1' ;
      Ha1a2 , Ha2    ,Ha2a3 , Hpa2    , Hba2' , Hwa2' , Hea2' ;
      Ha1a3 , Ha2a3 , Ha3    , Hpa3    , Hba3' , Hwa3' , Hea3' ;
      Hpa1   , Hpa2   , Hpa3   , Hphi   , Hphib', Hphiw0', Hphiw1';
      Hba1   , Hba2   , Hba3   , Hphib  , Hb    , Hbw0  , Hbw1  ;
      Hwa1   , Hwa2   , Hwa3   , Hphiw0 , Hbw0  , Hw0    , Hbw01  ;
      Hea1   , Hea2   , Hea3   , Hphiw1 , Hbw1  , Hbw01  , Hw1    ];

end
end
end

```

9.4.4 Simulación de valores de la distribución y recuperación de parámetros caso 0 inflacionada

```

clear all;
n = 1000; % simulated data size
% Define model parameters and independent variables%%%%%%%%
a =1.1; %new skew parameter
phi = 100;
betag = [0.1 0.2 -0.1 0.4 0.1]';
eta = [0.5 -1 -0.5 0.5 1]';
omega = [0.1 -0.1 -0.2 0.3 0.5]';
sigmas = [0.4 -0.5 0.4 0.2; -0.5 1 0.2 0.1; 0.4 0.2 3 1.5; 0.2 0.1 1.5
1]; %covarianza normal for independent variables
%sigmas = [0 0 0 0; 0 0 0 0; 0 0 0 0; 0 0 0 0];
kk = length(betag); % # of independent variables + intercept
X = mvnrnd(zeros(n,kk-1),sigmas);
X = [ones(n,1) X];
thetav = [a;a;a; phi ; betag ; omega ; eta];% true parameter
options = optimoptions('fmincon','Algorithm','trust-region-
reflective','TolX',1e-15,'TolFun', 1e-15,'MaxIter',10000, 'TolPCG', 0.0001,
'GradObj','on','Hessian','user-supplied','Display','off');
k = 3*kk+4;
nsim = 1; % number of simulations
res = [thetav', thetav', thetav',0,0];
param0 = zeros(k+1,1);
paramPW = zeros(k+1,1);
paramBIm = zeros(k+1,1);

```

```

warning('off', 'all');
yy = zeros(n,1);

for j=1:nsim,
try
    Y = sbeinfm_cloglog_01(X,thetav);
    index0 = find(Y==0);
    n0 = length(index0);
    index1 = find(Y==1);
    n1 = length(index1);
    indc = setdiff((1:n)',union(index0,index1));
    Xc = X(indc,:);
    Yc = Y(indc);
    m = n-n0-n1;
    Ym = Y+2;
% Model specification
    if m==0 % Binary regression model
        model = -1;
        Ym(Ym==3)=1;
        [w, dev] =
mnrfit(X(:,2:end),Ym,'model','nominal','interactions','on');
    elseif ((n0 > 0)&(n1 == 0)) % 0-inflated regression model
        model = 1;
        Ym(indc) = 1;
    elseif ((n0 == 0)&(n1 > 0)) %1-inflated regression model
        model = 2;
        Ym = Ym-1;
        Ym(indc) = 1;
    elseif ((n0 > 0)&(n1 > 0) & (m > 0)) %01-inflated regression model
        model = 3;
        Ym = Ym-1;
        Ym(indc) = 3;
    else % beta regression model
        model = 0;
    end

%----- Initial estimators-----
    initheta_cloglog01 %produces phi0, beta0, w0 and eta0
    beta0=beta1';
    w0=w1';
    eta0=eta1';

```

```

theta00 = [1,1,1,phi0,beta0,w0,eta0]';
%theta00=[0.900,0.900,0.900,105,0.678,0.798,-0.598,0.612,0.542,0.912,-
1.643,-1.443,0.523,0.534,1.123,-0.245,-0.212,0.367,0.545]';
%beta0=[0.678,0.798,-0.598,0.612,0.542];
%----- Papke-wooldridge methodology-----
tic;
[betaP,fval,exitflag] = fmincon('kfunpapke',beta0,[],[],[],[], -
1000*ones(1,kk), 1000*ones(1,kk),[], options,X,Y);
eflagP = exitflag;
%----- Two-part methodology-----
%maintwopart;
%eflag2 = exitflag;
%f2partc = -fval;
%theta00=[0.455,0.455,0.455,105.453,0.978,0.998,-
0.998,0.412,0.842,0.912,0.543,0.443,-0.523,-0.534,0.123,-0.245,-
0.212,0.367,0.545]';

%----- Our proposal-----
mainbinf;
eflag = exitflag;
fBIm = -fval;
%t3 = toc;
res = [res; theta0', [0;0;0;0;betaP';zeros(kk+5,1)]',
theta',eflag,fBIm];
yy = [yy,Y];

catch disp('Error'); %# Catch the exception
continue %# Pass control to the next loop iteration
end
end

par0 = res(2:end,1:k)';
parPW = res(2:end,k+1:2*k)';
parBIm = res(2:end,2*k+1:(end-2))';
mInitial = mean(par0,2);
mPW = mean(parPW,2);
mBIm = mean(parBIm,2);
[a,ssim] = size(par0);
RSMEInitial = sqrt(mean((par0 -kron(ones(1,ssim), thetav)).^2,2));
RSMEPW = sqrt(mean((parPW -kron(ones(1,ssim), thetav)).^2,2));
RSMEBIm = sqrt(mean((parBIm -kron(ones(1,ssim), thetav)).^2,2));
hist(Y,5);

```

```
table(thetav, mInitial, mPW, mBIM, RSMEInitial , RSMEPW, RSMEBIM)
```

9.4.5 Simulación de valores de la distribución y recuperación de parámetros caso 0-1

inflacionada

```
clear all;
n = 1000; % simulated data size
% Define model parameters and independent variables%%%%%%%%%
a = 0.5; %new skew parameter
phi = 20;
betag = [0.1 0.2 -0.1 0.4 0.1]';
%eta = [0.1 -0.1 -0.2 0.3 0.5]';
omega = [0.1 -0.1 -0.2 0.3 0.5]';
sigmas = [0.4 -0.5 0.4 0.2; -0.5 1 0.2 0.1; 0.4 0.2 3 1.5; 0.2 0.1 1.5
1]; %covarianza normal for independent variables
%sigmas = [0 0 0 0; 0 0 0 0; 0 0 0 0; 0 0 0 0];
kk = length(betag); % # of independent variables + intercept
X = mvnrnd(zeros(n, kk-1), sigmas);
X = [ones(n, 1) X];
%thetav = [a; a; a; phi ; betag ; omega ; eta]; % true parameter
thetav = [a; a; phi; betag; omega]; % true parameter
%thetav = [a; a; a; phi ; betag; eta]; % true parameter
options = optimoptions('fmincon', 'Algorithm', 'trust-region-
reflective', 'TolX', 1e-15, 'TolFun', 1e-15, 'MaxIter', 10000, 'TolPCG', 0.0001,
'GradObj', 'on', 'Hessian', 'user-supplied', 'Display', 'off');
k = 2*kk+3; %numero de parametros
nsim = 5; % number of simulations
res = [thetav', thetav', thetav', 0, 0];
param0 = zeros(k+1, 1);
paramPW = zeros(k+1, 1);
paramBIM = zeros(k+1, 1);
warning('off', 'all');
yy = zeros(n, 1);

for j=1:nsim,
try
Y = sbeinfm_cloglog_0(X, thetav);
index0 = find(Y==0);
n0 = length(index0);
index1 = find(Y==1);
```

```

n1 = length(index1);
indc = setdiff((1:n)',union(index0,index1));
Xc = X(indc,:);
Yc = Y(indc);
m = n-n0-n1;
Ym = Y+2;
% Model specification
if m==0 % Binary regression model
    model = -1;
    Ym(Ym==3)=1;
    [w, dev] =
mnrfit(X(:,2:end),Ym,'model','nominal','interactions','on');
elseif ((n0 > 0)&(n1 == 0)) % 0-inflated regression model
    model = 1;
    Ym(indc) = 1;
elseif ((n0 == 0)&(n1 > 0)) %1-inflated regression model
    model = 2;
    Ym = Ym-1;
    Ym(indc) = 1;
elseif ((n0 > 0)&(n1 > 0) & (m > 0)) %01-inflated regression model
    model = 3;
    Ym = Ym-1;
    Ym(indc) = 3;
else % beta regression model
    model = 0;
end
%----- Initial estimators-----
thetav(1)=0.5;
initheta_cloglog %produces phi0, beta0, w0 and eta0
beta0=beta1';
w0=w1';
%eta0=eta1';
theta00 = [0.5,0.5,phi0,beta0,w0]';
%theta00=[1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,1,-1,-1,1,1]';
%beta0=[1,1,-1,1,1];
%----- Papke-wooldridge methodology-----
tic;
[betaP,fval,exitflag] = fmincon('kfunpapke',beta0,[],[],[],[], -
1000*ones(1,kk), 1000*ones(1,kk),[], options,X,Y);
eflagP = exitflag;
%----- Two-part methodology-----

```

```

%maintwopart;
%eflag2 = exitflag;
%f2partc = -fval;

%----- Our proposal-----
mainbinf;
eflag = exitflag;
fBIm = -fval;
%t3 = toc;
res = [res; theta0', [0;0;0;betaP';zeros(kk,1)]', theta',eflag,fBIm];
yy = [yy,Y];
catch disp('Error'); %# Catch the exception
continue      %# Pass control to the next loop iteration
end
end
par0 = res(2:end,1:k)';
parPW = res(2:end,k+1:2*k)';
parBIm = res(2:end,2*k+1:(end-2))';
mInitial = mean(par0,2);
mPW = mean(parPW,2);
mBIm = mean(parBIm,2);
[a,ssim] = size(par0);
RSMEInitial = sqrt(mean((par0 -kron(ones(1,ssim), thetav)).^2,2));
RSMEPW = sqrt(mean((parPW -kron(ones(1,ssim), thetav)).^2,2));
RSMEBIm = sqrt(mean((parBIm -kron(ones(1,ssim), thetav)).^2,2));
hist(Y,5);
table(thetav, mInitial, mPW, mBIm, RSMEInitial , RSMEPW, RSMEBIm)

```