



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Tests de Wald Robustos en el
Modelo de Regresión Log Gamma

Jimena Soledad Blaiotta

Directora: Dra. Ana M. Bianco

11 de noviembre de 2010

Índice general

1. Introducción	3
2. Robustez	11
2.1. Modelo de posición y escala	12
2.1.1. Modelo de posición	12
2.1.2. M -estimadores de posición	13
2.1.3. Estimadores de dispersión	16
2.1.4. M -estimadores de escala	17
2.1.5. M -estimadores de posición con escala desconocida	18
2.2. Modelo de regresión lineal con predictores fijos	19
2.2.1. M -estimadores	20
2.3. Modelo de regresión lineal con predictores aleatorios	22
2.3.1. MM -estimadores	22
2.3.2. S -estimadores	24
3. Estimación en el Modelo de Regresión Log–Gamma	25
3.1. M -estimadores	26
3.2. S -estimadores	30
3.3. MM -estimadores	33
3.4. Aspectos computacionales	35
4. Test de Tipo Wald en el Modelo de Regresión Log–Gamma	39
4.1. Estadístico de Wald	39
4.2. Comportamiento asintótico	40
4.2.1. Estadístico clásico	41
4.2.2. Estadístico robusto	43
5. Simulaciones	51
5.1. Nivel	51
5.2. Potencia	58
6. Ejemplos	65
6.1. Datos de Feigl y Zelen	65
6.2. Enfermedades Vasculares	68

Agradecimientos

A mi mamá, porque todo lo que soy te lo debo a vos.

Porque me apoyas y acompañas en todas mis decisiones, aun en la decisión de estudiar, a pesar de que no era un camino fácil y natural. Este también es tu logro.

Por tus consejos, tu paciencia y por guiarme siempre.

A Pablo, por ser mi compañero incondicional, por ayudarme, alentarme y sostenerme siempre. Por todas las alegrías, caídas y logros que vivimos juntos.

A mis hermanos y mis abuelos, por entender mis ausencias debidas al estudio, por su cariño y apoyo.

A Bety, por estar siempre. Por todo lo que me ayudaste en el momento de elegir que estudiar y por los granitos de arena que hicieron que los comienzos de esta etapa no sean tan duros.

A la tía Ana, por brindarme tu casa con toda generosidad, y hacer muchas veces que mis viajes a la facultad no sean interminables.

A mis amigas, Lau, Georgi y Mariela, por todos los nervios, alegrías y esfuerzos que compartimos en distintas etapas a lo largo de la carrera.

Por la amistad que tenemos.

A Susana Puddu y Fabio Vicentini, porque gracias a ustedes descubrí el placer por la matemática aplicada.

A mi directora, Ana, porque dentro de la matemática aplicada, me hiciste sentir el gusto por la Estadística.

Por la paciencia que me tuviste a lo largo del desarrollo de esta tesis, por entender mis tiempos, y por todo el esfuerzo que dedicaste a la realización de este trabajo.

A todos los que me acompañaron a transitar este camino

gracias!

Capítulo 1

Introducción

En muchas oportunidades un investigador se interesa en conocer cómo el comportamiento de un conjunto de variables impacta sobre otra, entender y describir la posible relación funcional entre las mismas y a partir de esto, formular un modelo de predicción.

Por ejemplo, supongamos que se observan n vectores aleatorios de la forma (\mathbf{x}'_i, y_i) $1 \leq i \leq n$, donde $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$, es el i -ésimo vector de variables explicativas (predictores o independientes) e y_i son variables de respuesta (o dependientes) y se intenta determinar la relación que existe entre ellas a través de una forma funcional f . Una función natural y simple es la forma lineal, es decir, aquella en que la respuesta y es una combinación lineal de las p variables explicativas, o sea planteamos

$$y_i = \sum_{j=1}^p \mathbf{x}_{ij} \beta_j + u_i, \quad i = 1, \dots, n, \quad (1.1)$$

donde $\beta' = (\beta_1, \dots, \beta_p)$ son los parámetros desconocidos a ser estimados y u_i 's representan los errores que son variables aleatorias independientes e idénticamente distribuidas (i.i.d.) e independientes de \mathbf{x}_i . Si el modelo tiene intercept u ordenada al origen, la última componente de \mathbf{x} será un 1. En general, asumimos para todo $i = 1, \dots, n$

ML1. $E(u_i) = 0$

ML2. $\text{COV}(u_i, u_j) = 0$ si $i \neq j$, y $\text{VAR}(u_i) = \sigma^2$,

lo que indica que no existe un aporte sistemático de los errores en el modelo y se garantiza la ausencia de correlación y la homoscedasticidad en los errores. Este modelo, que es conocido como *Modelo Lineal*, es extensamente usado, entre otras razones, por la facilidad de su interpretación ya que al ser aditivo es fácilmente comprensible el efecto de una variable explicativa sobre la respuesta.

Este modelo ha sido ampliamente estudiado y las propiedades de los estimadores de mínimos cuadrados son bien conocidas bajo ciertas condiciones de regularidad. Si además, para todo $i = 1, \dots, n$, se cumple que

ML3. $u_i \sim N(0, \sigma^2)$,

el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ coincide con el de mínimos cuadrados y su distribución exacta es conocida, lo que es de gran utilidad en la etapa de inferencia cuando se desean realizar tests de hipótesis o intervalos de confianza.

Si bien este modelo es muy sencillo, no siempre se ajusta a los datos. En términos de la esperanza de las variables de respuesta, bajo la hipótesis **ML1** el modelo (1.1) puede reescribirse como

$$E(y_i) = \sum_{j=1}^p x_{ij}\beta_j = \mathbf{x}'_i\boldsymbol{\beta} \quad i = 1, \dots, n.$$

Sin embargo, es claro que este supuesto no podrá verificarse para variables de respuesta tales como Poisson o Gamma, en las que a priori la esperanza está limitada a tomar valores positivos, mientras que los predictores lineales $\mathbf{x}'_i\boldsymbol{\beta}$ no están restringidos, a menos que se impongan restricciones adicionales. Una flexibilización de este modelo que permite tratar situaciones como las mencionadas, es la generalización que se logra con el *Modelo Lineal Generalizado* (GLM).

En el modelo lineal generalizado, se relaciona la distribución de la variable de respuesta y con la componente sistemática, es decir las covariables \mathbf{x} , a través de una función de enlace $H(x)$ que relaciona la esperanza de las respuestas con el predictor lineal, es decir

$$H(E(y_i)) = \mathbf{x}'_i\boldsymbol{\beta} \quad i = 1, \dots, n.$$

El Modelo Lineal Generalizado fue introducido por Nelder y Wedderburn (1974) como una manera de unificar varios modelos estadísticos, incluyendo la regresión lineal, regresión logística y regresión de Poisson, bajo un solo marco teórico. En esta tesis, trabajaremos con un caso particular, que es el de la regresión Gamma.

Más precisamente, consideremos la parametrización de una función de distribución Gamma, denotada por $\Gamma(\alpha, \mu)$, con $\alpha > 0$ y $\mu > 0$, con función de densidad dada por

$$f(z, \alpha, \mu) = \begin{cases} \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} z^{\alpha-1} \exp(-(\frac{\alpha}{\mu})z) & \text{si } z \geq 0 \\ 0 & \text{si } z < 0. \end{cases}$$

De esta forma, si z es una variable aleatoria con distribución $\Gamma(\alpha, \mu)$, α es un parámetro de forma, $E(z) = \mu$ y $\text{VAR}(z) = \frac{\mu^2}{\alpha}$.

Consideremos el Modelo Lineal Generalizado donde la distribución de la respuesta z dado el vector de covariables $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ es $\Gamma(\alpha, \mu(\mathbf{x}))$, de tal forma que

$$z|\mathbf{x} \sim \Gamma(\alpha, \mu(\mathbf{x})),$$

donde $\log(\mu(\mathbf{x})) = \mathbf{x}'\boldsymbol{\beta}$ y por lo tanto, $H(t) = \log(t)$ es la función de enlace o link, que relaciona la media de la variable de respuesta con el predictor lineal basado en las variables explicativas. Asumiremos que el parámetro de forma α no depende de \mathbf{x} .

Este modelo también puede verse como un modelo lineal con errores asimétricos. De hecho, si llamamos $\epsilon = \frac{z}{\mu(x)}$, esta nueva variable aleatoria tiene distribución $\Gamma(\alpha, 1)$, luego definiendo $y = \log(z)$ y $u = \log(\epsilon)$ este modelo puede ser reescrito de la forma

$$y = \mathbf{x}'\boldsymbol{\beta} + u, \quad (1.2)$$

donde el error u tiene como función de distribución al logaritmo de $\Gamma(\alpha, 1)$, $\log \Gamma(\alpha, 1)$, \mathbf{x} y u son independientes y la función de densidad de u es $g(u, \alpha)$, con

$$g(u, \alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \exp(\alpha(u - \exp(u))). \quad (1.3)$$

Esta densidad es asimétrica y unimodal, con máximo en 0. El modelo dado en (1.2) es llamado *Modelo de Regresión Log-Gamma* (LGR).

Es interesante notar que la densidad (1.3) pertenece a una familia exponencial de la forma

$$g(u, \alpha) = Q(\alpha) \exp(\alpha t(u)), \quad (1.4)$$

con $t(u) = u - \exp(u)$. Las distribuciones de esta forma son continuas y con una única moda en u_0 .

Si $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ es una muestra aleatoria que cumple el modelo (1.2) con errores u_i con densidad de la forma (1.4), el estimador de máxima verosimilitud de α y de $\boldsymbol{\beta}$ son los valores que maximizan la verosimilitud $L(a, \mathbf{b})$ siendo

$$L(a, \mathbf{b}) = \prod_{i=1}^n g(y_i - \mathbf{x}'_i \mathbf{b}, a).$$

Como veremos en el Capítulo 3, la maximización de $L(a, \mathbf{b})$ respecto de \mathbf{b} es equivalente a la minimización de la *deviance*, que compara el logaritmo de la verosimilitud bajo el modelo saturado con el logaritmo de la verosimilitud del modelo que estamos considerando y puede escribirse como

$$D(a, \mathbf{b}) = \sum_{i=1}^n 2 \log \left(\frac{g(u_0, a)}{g(y_i - \mathbf{x}'_i \mathbf{b}, a)} \right). \quad (1.5)$$

Si llamamos $d_i(a, \mathbf{b})$ a la i -ésima componente de la deviance, entonces

$$d_i(a, \mathbf{b}) = 2 \log \left(\frac{g(u_0, a)}{g(y_i - \mathbf{x}'_i \mathbf{b}, a)} \right)$$

Es muy sencillo ver que

$$d_i(a, \mathbf{b}) = 2a(t(u_0) - t(y_i - \mathbf{x}'_i \mathbf{b})) = 2a d^*(y_i, \mathbf{x}_i, \mathbf{b}),$$

donde

$$d^*(y, \mathbf{x}, \mathbf{b}) = t(u_0) - t(y - \mathbf{x}' \mathbf{b}). \quad (1.6)$$

En el caso particular que nos interesa de la distribución log-Gamma tenemos que $t(u) = u - \exp(u)$ y por lo tanto d^* resulta

$$d^*(y, \mathbf{x}, \mathbf{b}) = -1 - (y - \mathbf{x}'\mathbf{b}) + \exp(y - \mathbf{x}'\mathbf{b}).$$

Luego, el estimador de máxima verosimilitud de β lo podemos obtener como

$$\widehat{\beta}_n^{ML} = \arg \min_{\mathbf{b}} \sum_{i=1}^n d^*(y_i, \mathbf{x}_i, \mathbf{b}).$$

Sin embargo, es bien conocido que en el contexto de los modelos de regresión los estimadores de mínimos cuadrados y de máxima verosimilitud pueden ser severamente afectados por la presencia de datos atípicos. Entre las numerosas propuestas de estimadores robustos en el modelo de regresión lineal se destacan los *MM*-estimadores propuestos por Yohai (1987) que tienen la propiedad de poseer un alto punto de ruptura, a la vez que conservan una alta eficiencia. Como otras propuestas realizadas, ésta tiene la desventaja de que cuando la distribución de los errores es asimétrica, tal como es el caso de la distribución log-Gamma, si bien los estimadores de la pendiente son consistentes, los estimadores de la ordenada al origen dejan de serlo. En consecuencia, también pueden resultar no consistentes las estimaciones de funciones paramétricas tales como las predicciones de la media condicional.

En el caso de los modelos lineales generalizados, los estimadores de máxima verosimilitud son muy sensibles a la presencia de datos anómalos y numerosas propuestas robustas han sido realizadas. En particular, para la estimación de los parámetros de una distribución Gamma, Marazzi y Ruffieux (1996) consideraron el caso de respuestas i.i.d., Stefanski, Carroll y Ruppert (1986) y Künsch, Stefanski y Carroll (1989) obtuvieron estimadores de influencia acotada óptimos en el sentido de Hampel para modelos lineales generalizados. Más tarde, Cantoni y Ronchetti (2001) introdujeron estimadores robustos basados en el concepto de la cuasi-verosimilitud.

Una desventaja de todos estos estimadores es que requieren el cómputo de un término de corrección a fin de obtener estimaciones Fisher-consistentes.

Bianco, García Ben y Yohai (2005) generalizan los *MM*-estimadores al caso en que los errores tienen una distribución en la clase de las familias exponenciales del tipo (1.4), que como vimos incluye la distribución log-Gamma. Como veremos, la definición de estos estimadores es muy parecida a la de los estimadores originales introducidos por Yohai (1987) con la diferencia de que los residuos ordinarios son reemplazados por los residuos basados en la deviance. Estos autores muestran que los estimadores propuestos son Fisher-consistentes, derivan su distribución asintótica, calculan su punto de ruptura y muestran que poseen al mismo tiempo alto punto de ruptura y alta eficiencia. Una ventaja de esta generalización de los *MM*-estimadores es que no necesitan un término de corrección para obtener Fisher-consistencia bajo ciertas condiciones de regularidad.

El interés del presente trabajo está centrado en proponer tests de hipótesis robustos en el modelo de regresión log-Gamma, problema que ha recibido mucho menos atención que el de estimación robusta. Sería deseable obtener tests de hipótesis que conserven un nivel estable y mantengan una buena potencia bajo contaminación. En efecto, los métodos de inferencia, tales como los tests y los intervalos de confianza, heredan de los procedimientos de estimación en los que se basan su sensibilidad a la presencia de datos atípicos. En este sentido una pequeña proporción de datos atípicos puede afectar severamente tanto el nivel como la potencia de los tests derivados. Los trabajos de Heritier y Ronchetti (1994) y Cantoni y Ronchetti (2001) proponen métodos en esta dirección. Heritier y Ronchetti (1994) proponen tests robustos para un modelo paramétrico general. Por otro lado, Cantoni y Ronchetti (2001), usando la noción de cuasi-verosimilitud, definen versiones robustas de la deviance e introducen una familia de tests para el problema de selección de variables en modelos lineales generalizados. Asimismo, investigan la estabilidad del nivel asintótico de los procedimientos propuestos.

En esta tesis se propone estudiar algunas propiedades asintóticas de tests de tipo Wald basados en los MM -estimadores introducidos en Bianco, García Ben y Yohai (2005) y se muestra que bajo la hipótesis nula tiene el mismo comportamiento asintótico que su equivalente clásico basado en los estimadores de máxima verosimilitud. Mediante un estudio de Monte Carlo se investiga su comportamiento para muestras finitas bajo el modelo y bajo contaminación.

La organización de la tesis se describe a continuación. En el Capítulo 2 se introducen algunas nociones de robustez y algunos de los estimadores robustos más difundidos para el modelo lineal. En el Capítulo 3 se hace una revisión de los MM -estimadores introducidos en Bianco, García Ben y Yohai (2005) y de sus propiedades, así como de algunos aspectos computacionales. En el Capítulo 4 se presenta el estadístico del test y se derivan sus propiedades asintóticas. Finalmente, en el Capítulo 5 se presentan los resultados de un estudio de Monte Carlo y en el Capítulo 6 se analizan dos ejemplos de datos reales.

Capítulo 2

Robustez

En Estadística, los métodos clásicos se basan en supuestos fuertes sobre la distribución de los datos, tales como la suposición de normalidad. Los procedimientos que se obtienen bajo estas hipótesis son óptimos cuando estas suposiciones se cumplen exactamente, pero pueden ser altamente sensibles y muy ineficientes frente a cualquier desviación de las mismas. Una pequeña proporción de datos atípicos puede alterar las conclusiones obtenidas con estos procedimientos.

Es muy frecuente que en los datos analizados haya unas pocas observaciones atípicas que pueden deberse a distintas razones, como por ejemplo: a errores en los instrumentos de medición, a la variación de las condiciones bajo las cuales se obtuvieron los datos, etc. Cuando se usan los métodos clásicos, unos pocos datos atípicos, inclusive uno solo, pueden tener un efecto muy importante en los resultados del análisis estadístico e invalidar las conclusiones. Por ejemplo, cuando se ajusta un modelo de regresión lineal, el procedimiento óptimo es el de mínimos cuadrados, que es muy sensible a datos atípicos.

La primera salida frente a este problema, fue el desarrollo de procedimientos para detectar estos datos atípicos y suprimirlos del análisis. Una vez suprimidos, al resto de los datos se aplicaban los procedimientos clásicos.

El enfoque de la estadística robusta es distinto, ya que reemplaza la detección y supresión de datos atípicos por el desarrollo de procedimientos que son muy poco afectados por la presencia de éstos.

En este capítulo mostraremos como conciliar robustez y eficiencia, utilizando estimadores que simultáneamente sean altamente robustos cuando las observaciones tienen datos atípicos y altamente eficientes cuando los errores son normales sin contaminación.

Maronna, Martin y Yohai (2006) dan una revisión muy completa sobre los temas que desarrollamos en este capítulo.

2.1. Modelo de posición y escala

2.1.1. Modelo de posición

Supongamos que tenemos n observaciones de variables x_1, \dots, x_n , que dependen de un parámetro desconocido μ y de ciertos errores aleatorios. Una expresión matemática simple que describe este proceso, es la que resulta de asumir que dichos errores son aditivos, es decir:

$$x_i = \mu + u_i, 1 \leq i \leq n, \quad (2.1)$$

donde los errores, u_1, \dots, u_n , son variables aleatorias. Este modelo es llamado *modelo de posición*. Si las observaciones son el resultado de repeticiones independientes del mismo experimento, podemos asumir que

- u_1, \dots, u_n tienen la misma función de distribución F_0 , y son independientes.

Resulta entonces que x_1, \dots, x_n son variables aleatorias independientes e idénticamente distribuidas (i.i.d.), con función de distribución

$$F(x) = F_0(x - \mu). \quad (2.2)$$

Un estimador $\hat{\mu}_n$ es una función de las observaciones $\hat{\mu}_n = \hat{\mu}(x_1, \dots, x_n) = \hat{\mu}(\mathbf{x})$. Buscamos estimadores que en algún sentido hagan que $\hat{\mu}_n \approx \mu$ con alta probabilidad. Una forma de medir esta aproximación es con el *Error Cuadrático Medio* (MSE)

$$MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2.$$

Notemos que si $\hat{\mu}$ es la media muestral y los datos son normales, este estimador es “óptimo”, es decir, coincide con el estimador de máxima verosimilitud (MLE) y minimiza el MSE sobre los estimadores insesgados.

Sin embargo, en la mayoría de la aplicaciones prácticas a lo sumo se puede asegurar que los errores de medición tienen distribución *aproximadamente normal*. Una forma de caracterizar distribuciones *aproximadamente normales* es considerar entornos de contaminación de tamaño ϵ de F , decimos que F es una distribución normal contaminada si

$$F = (1 - \epsilon)G + \epsilon H, \quad (2.3)$$

donde $G = N(\mu, \sigma^2)$ y H puede ser cualquier distribución.

Se puede ver que aunque ϵ sea pequeño, la media muestral puede ser muy ineficiente, por este motivo, el objetivo es buscar estimadores casi tan “buenos” como lo es la media muestral cuando la distribución es exactamente normal, pero que también sean “buenos” en algún sentido cuando F es aproximadamente normal.

Si tomamos la mediana como estimador, se puede ver que el beneficio de la robustez obtenido se paga con una pérdida en eficiencia bajo la distribución normal. Por esta razón, se han desarrollado estimadores que combinen la alta eficiencia de la media bajo normalidad con la robustez de la mediana bajo contaminación.

2.1.2. M -estimadores de posición

Describiremos una familia de estimadores que contienen a la media y a la mediana como casos especiales.

Consideremos el modelo (2.1) y asumamos que F_0 tiene función de densidad $f_0 = F_0'$. Luego, la densidad conjunta de las observaciones (*función de verosimilitud*) es

$$L(x_1, \dots, x_n; t) = \prod_{i=1}^n f_0(x_i - t).$$

El estimador de máxima verosimilitud (MLE) de μ es aquel valor que maximiza la función de verosimilitud, es decir

$$\hat{\mu}_n = \arg \max_t L(x_1, \dots, x_n; t).$$

Si f_0 es positiva,

$$\hat{\mu}_n = \arg \min_t \sum_{i=1}^n \rho(x_i - t), \quad (2.4)$$

con

$$\rho = -\log f_0.$$

Si ρ (función de pérdida) es derivable, derivando (2.4) con respecto a t , obtenemos

$$\sum_{i=1}^n \psi(x_i - \hat{\mu}_n) = 0, \quad (2.5)$$

con $\psi = \rho'$ (función de score).

Observemos que para $\rho(x) = x^2/2$ obtenemos $\hat{\mu}_n = \bar{x}$ como solución de (2.4) y en el caso de $\rho = |x|$ se puede ver que cualquier mediana será solución de esa ecuación.

Dada una función de pérdida ρ , un *M-estimador de posición* es una solución de (2.4). Nosotros estudiaremos estimadores de esta forma, que no necesariamente tienen que ser MLE para alguna distribución.

Si ψ es monótona no decreciente y continua, con $\psi(-\infty) < 0 < \psi(\infty)$, se puede probar que (2.4) y (2.5) siempre tienen solución. Si además, ψ estrictamente creciente, la solución es única.

Es fácil ver que los M -estimadores son equivariantes por traslaciones ($\hat{\mu}(x_1 + c, \dots, x_n + c) = \hat{\mu}(x_1, \dots, x_n) + c$), aunque en general no son equivariantes por cambios de escala ($\hat{\mu}(cx_1, \dots, cx_n) = |c|\hat{\mu}(x_1, \dots, x_n)$), como lo son la media y la mediana.

Para evaluar el rendimiento de los M -estimadores, es necesario calcular su distribución. Asumiendo que ψ es creciente, definimos $\mu = \mu(F)$ como la solución de

$$E_F \psi(x - t) = 0.$$

Se puede ver que el M -estimador $\hat{\mu}_n$ es consistente a μ , su *distribución aproximada* es $N\left(\mu, \frac{v}{n}\right)$, y la *varianza asintótica* es

$$v = \frac{E_{F_0}(\psi(x - \mu)^2)}{(E_{F_0}\psi'(x - \mu))^2}.$$

Notemos que bajo el modelo (2.2) v no depende de μ , es decir

$$v = \frac{E_{F_0}(\psi(x)^2)}{(E_{F_0}\psi'(x))^2}.$$

Definimos la *eficiencia asintótica* de $\hat{\mu}_n$ como

$$Eff(\hat{\mu}) = \frac{v_{MLE}}{v},$$

siendo v_{MLE} la varianza asintótica del estimador de máxima verosimilitud.

A modo de ejemplo, un tipo de funciones ρ de particular importancia es la familia de *funciones de Huber*:

$$\rho(x) = \begin{cases} x^2 & \text{si } |x| \leq k \\ 2k|x| - k^2 & \text{si } |x| > k. \end{cases}$$

Los M -estimadores que corresponden a los casos límite $k \rightarrow \infty$ y $k \rightarrow 0$ son la media y la mediana, respectivamente. El valor de k es elegido en función de la eficiencia asintótica deseada. En general, se consideran funciones de pérdida que satisfacen la siguiente definición.

Definición 2.1 Decimos que una función es una ρ -función, si verifica:

R1. $\rho(x)$ es una función no decreciente de $|x|$.

R2. $\rho(0) = 0$

R3. $\rho(x)$ creciente para $x > 0$ y tal que $\rho(x) < \rho(\infty)$

R4. Si ρ es una función acotada, entonces, $\rho(\infty) = 1$

M -estimadores redescendientes

Un tipo muy usado de funciones ρ son las pertenecientes a la familia bicuadrada dada por

$$\rho(x) = \begin{cases} 1 - [1 - (\frac{x}{k})^2]^3 & \text{si } |x| \leq k \\ 1 & \text{si } |x| > k \end{cases} \quad (2.6)$$

Los M -estimadores que resultan de usar una función ρ bicuadrada, serán llamados “ M -estimadores redescendientes” y resultan ser más resistentes ante la presencia de grandes datos atípicos.

2.1.3. Estimadores de dispersión

La forma tradicional de medir la variabilidad de un conjunto de datos $\mathbf{x} = [x_1, \dots, x_n]$ es con el desvío estándar (SD)

$$SD(\mathbf{x}) = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}.$$

Para cualquier constante c el SD satisface las condiciones de invariancia por traslaciones y equivariancia por cambios de escala

$$SD(\mathbf{x} + c) = SD(\mathbf{x}), \quad SD(c\mathbf{x}) = |c|SD(\mathbf{x}). \quad (2.7)$$

Un estadístico que satisface las condiciones (2.7) se llama *estimador de dispersión*.

Como el desvío standard SD es muy poco robusto, una alternativa puede ser la *media de desviaciones absolutas* (MD)

$$MD(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|,$$

que también es sensible a outliers, pero lo es menos que el SD . Podemos notar que ambos estadísticos utilizan la media (esto es necesario para lograr la invariancia por traslaciones), pero por este motivo son tan sensibles a datos atípicos.

Una alternativa robusta es utilizar la mediana en lugar de la media, resultando la *mediana de desviaciones absolutas* (MAD)

$$MAD(\mathbf{x}) = \text{Med}(|\mathbf{x} - \text{Med}(\mathbf{x})|).$$

2.1.4. M -estimadores de escala

Consideremos el modelo multiplicativo

$$x_i = \sigma u_i, \quad (2.8)$$

donde u_1, \dots, u_n son variables i.i.d con densidad f_0 y $\sigma > 0$ es el parámetro desconocido.

La función de densidad de cada x_i será

$$\frac{1}{\sigma} f_0\left(\frac{x}{\sigma}\right).$$

El estimador de máxima verosimilitud de σ en (2.8) es

$$\hat{\sigma}_n = \arg \max_s \frac{1}{s^n} \prod_{i=1}^n f_0\left(\frac{x_i}{s}\right).$$

Tomando logaritmo y derivando respecto a s , obtenemos

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{s}\right) = 1,$$

donde $\rho(t) = t\psi(t)$ con $\psi = -f'_0/f_0$.

En general, llamaremos *M-estimador de escala* a aquel estimador que satisfaga una ecuación de la forma:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{x_i}{s}\right) = \delta, \quad (2.9)$$

donde ρ es una ρ -función y δ es una constante positiva.

Es fácil verificar que un *M-estimador* es equivariante por cambio de escala, es decir $\hat{\sigma}(c\mathbf{x}) = c\hat{\sigma}(\mathbf{x}) \forall c > 0$, y si ρ es par $\hat{\sigma}(c\mathbf{x}) = |c|\hat{\sigma}(\mathbf{x}) \forall c$.

Para valores de n grande, la sucesión de estimadores (2.9) converge a

$$E_\rho\left(\frac{x}{\sigma}\right) = \delta.$$

2.1.5. *M-estimadores de posición con escala desconocida*

Como aclaramos previamente, en general los estimadores definidos en (2.4) no son equivariantes por cambio de escala, lo cual implica que los resultados pueden depender de las unidades de medida.

Para fijar ideas, supongamos que queremos estimar μ en el modelo (2.1) donde F está dada por (2.3), con $G = N(\mu, \sigma^2)$.

Si σ fuera conocido, sería natural dividir por σ en (2.1) y reducir el problema al caso $\sigma = 1$, lo cual implica estimar μ por

$$\hat{\mu}_n = \arg \min_t \sum_{i=1}^n \rho\left(\frac{x_i - t}{\sigma}\right).$$

Es fácil verificar que la distribución aproximada de $\hat{\mu}_n$ así definido es $N(\mu, v/n)$ con

$$v = \sigma^2 \frac{E(\psi((x - \mu)/\sigma)^2)}{(E(\psi'((x - \mu)/\sigma)))^2}.$$

Para obtener un *M-estimador de posición*, que sea equivariante por cambio de escala, una idea intuitiva es usar

$$\hat{\mu}_n = \arg \min_t \sum_{i=1}^n \rho\left(\frac{x_i - t}{\hat{\sigma}_n}\right), \quad (2.10)$$

donde $\hat{\sigma}_n$ es un estimador de dispersión computado previamente. Dado que $\hat{\sigma}_n$ estima la escala de los residuos, es intuitiva la necesidad de que sea robusto. Es fácil verificar que $\hat{\mu}_n$ así definido resulta equivariante por cambio de escala. Dado que $\hat{\sigma}_n$ no depende de μ , (2.10) implica que $\hat{\mu}_n$ satisface

$$\sum_{i=1}^n \psi \left(\frac{x_i - \hat{\mu}_n}{\hat{\sigma}_n} \right) = 0,$$

Una alternativa a esta forma de estimación, es considerar un modelo de posición-dispersión, con dos parámetros desconocidos y estimarlos simultáneamente (detalles de esto pueden verse en Maronna, Martin y Yohai, 2006), pero en general, la estimación con cómputo previo del parámetro de dispersión, suele ser más robusta que la estimación simultánea.

2.2. Modelo de regresión lineal con predictores fijos

Supongamos que se tienen n observaciones de la forma $(x_{i1}, \dots, x_{ip}, y_i)$, donde x_{i1}, \dots, x_{ip} son variables predictoras (covariables o variables independientes) e y_i es una variable de respuesta (o variable dependiente). Asumimos que se cumple el siguiente *modelo lineal*

$$y_i = \sum_{j=1}^p x_{ij} \beta_j + u_i, \quad i = 1, \dots, n,$$

donde β_1, \dots, β_p son los parámetros desconocidos a ser estimados y u_i 's son variables aleatorias i.i.d (los “errores”), que no dependen de \mathbf{x}_i 's.

En esta sección consideramos a (x_{i1}, \dots, x_{ip}) no aleatorios, fijos, es decir, determinados antes de realizar el experimento.

Llamando \mathbf{x}_i y $\boldsymbol{\beta}$ al vector columna p -dimensional con coordenadas $(x_{i1}, \dots, x_{ip})'$ y $(\beta_1, \dots, \beta_p)'$, respectivamente, el modelo puede ser escrito de la siguiente forma

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i. \quad (2.11)$$

Si llamamos \mathbf{X} a la matriz de $n \times p$ con elementos x_{ij} e \mathbf{y} y \mathbf{u} a los vectores con elementos y_i y u_i , respectivamente, el modelo lineal puede ser escrito en forma matricial como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \quad (2.12)$$

2.2.1. M -estimadores

Supongamos que se cumple el modelo (2.12), donde u_i tiene densidad

$$\frac{1}{\sigma} f_0 \left(\frac{u}{\sigma} \right)$$

siendo σ el parámetro de escala. En el modelo lineal (2.12), las y_i 's son independientes pero no idénticamente distribuidas, de hecho cada y_i tiene densidad

$$\frac{1}{\sigma} f_0 \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}}{\sigma} \right)$$

y por lo tanto, la función de verosimilitud para $\boldsymbol{\beta}$, asumiendo fijo el valor de σ es

$$L(\mathbf{b}) = \frac{1}{\sigma^n} \prod_{i=1}^n f_0 \left(\frac{y_i - \mathbf{x}'_i \mathbf{b}}{\sigma} \right).$$

Calcular el MLE maximizando $L(\mathbf{b})$ es equivalente a encontrar $\widehat{\boldsymbol{\beta}}_n$ tal que

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\mathbf{b}} \frac{1}{n} \sum_{i=1}^n \rho_0 \left(\frac{r_i(\mathbf{b})}{\sigma} \right) + \log \sigma,$$

donde $\rho_0 = -\log f_0$. Derivando respecto a \mathbf{b} obtenemos la ecuación equivalente

$$\sum_{i=1}^n \psi_0 \left(\frac{r_i(\mathbf{b})}{\sigma} \right) \mathbf{x}_i = 0,$$

donde $\psi_0 = \rho'_0 = -f'_0/f_0$.

Dada una función de pérdida ρ que satisface la Definición 2.1, definimos un *M-estimador de regresión* como

$$\widehat{\boldsymbol{\beta}}_n = \arg \min_{\mathbf{b}} \sum_{i=1}^n \rho \left(\frac{r_i(\mathbf{b})}{\widehat{\sigma}_n} \right), \quad (2.13)$$

donde $\widehat{\sigma}_n$ es un estimador de escala.

Derivando respecto a \mathbf{b} , obtenemos

$$\sum_{i=1}^n \psi \left(\frac{r_i(\mathbf{b})}{\widehat{\sigma}_n} \right) \mathbf{x}_i = 0, \quad (2.14)$$

donde $\psi = \rho'$. Esta última ecuación, no tiene que ser necesariamente la ecuación de un MLE.

Asumimos que la matriz \mathbf{X} tiene rango completo. En el caso particular en que σ es conocido, se puede verificar que el *M-estimador* es equivariante por regresión, traslaciones y cambios de escala.

Las soluciones de (2.14) con ψ monótona (respectivamente redescendiente), son llamadas *M-estimadores monótonos de regresión* (respectivamente *redescendientes*).

Las soluciones de (2.14) son soluciones de (2.13), y más aún, si ψ es estrictamente creciente, la solución es única. En el caso de regresión, al igual que para los *M-estimadores de posición redescendientes*, estos, a diferencia de los *M-estimadores monótonos* tienen un mejor balance entre eficiencia y robustez. Por este motivo, en general, utilizaremos un *M-estimador monótono* como punto inicial necesario para computar un *M-estimador redescendiente*.

Asumiendo que se cumple el modelo (2.11), con u tal que

$$E_{\psi} \left(\frac{u}{\sigma} \right) = 0 \quad (2.15)$$

y bajo condiciones de regularidad, $\widehat{\boldsymbol{\beta}}_n$ es consistente a $\boldsymbol{\beta}$ y además

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} N_p(0, v(\mathbf{X}'\mathbf{X})^{-1})$$

donde v esta dado por

$$v = \sigma^2 \frac{E(\psi((u/\sigma)^2))}{(E(\psi'(u/\sigma)))^2}. \quad (2.16)$$

Una demostración general esta dada por Yohai y Marona (1979).

2.3. Modelo de regresión lineal con predictores aleatorios

En esta sección consideramos n vectores aleatorios i.i.d. (\mathbf{x}'_i, y_i) de dimensión p , que satisfacen el modelo

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i.$$

Las u_i 's son i.i.d. e independientes de las \mathbf{x}'_i 's.

El análogo de asumir que X tiene rango completo, es asumir que la distribución de las \mathbf{x} no se concentra en ningún subespacio, es decir, $P(\mathbf{a}'\mathbf{x} = 0) < 1 \forall \mathbf{a} \neq 0$.

Bajo estas condiciones, suponiendo que las \mathbf{x}'_i 's tienen varianza finita, que $\hat{\sigma}_n$ es consistente a σ , y que se cumple (2.15), se puede probar que el estimador $\hat{\boldsymbol{\beta}}$ definido en (2.13) es consistente y asintóticamente normal, más aún

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} N_p(0, vV_x^{-1}),$$

donde $V_{\mathbf{x}} = E(\mathbf{x}\mathbf{x}')$ y \mathbf{x} tiene la misma distribución que las covariables \mathbf{x}_i y v fue definido en (2.16).

2.3.1. MM-estimadores

Computar un M -estimador requiere encontrar el mínimo absoluto de

$$L(\mathbf{b}) = \sum_{i=1}^n \rho \left(\frac{r_i(\mathbf{b})}{\hat{\sigma}_n} \right). \quad (2.17)$$

Cuando ρ es acotada suele ser una tarea sumamente difícil, excepto para $p = 1$ ó 2 . Sin embargo, se puede probar que es suficiente encontrar un “buen” mínimo local para alcanzar simultáneamente alto

punto de ruptura y alta eficiencia bajo la distribución normal. Este mínimo local se obtiene a partir de un punto de partida fiable y de aplicar el algoritmo iterativo de mínimos cuadrados reponderados (IRWLS).

Este estimador inicial también se utilizará para calcular un estimador de escala robusto, necesario para definir el M -estimador, y es necesario que éste se pueda calcular sin necesidad de una escala previa. Los estimadores que tienen estas características, son llamados S -estimadores, y los estudiaremos en la próxima sección.

Yohai (1987) definió un MM -estimador, como el resultado de los siguientes pasos:

1. Computar un estimador inicial $\hat{\beta}_0$ consistente, con alto punto de ruptura, aunque posiblemente de baja eficiencia.
2. Computar una escala robusta $\hat{\sigma}_n$ de los residuos $r_i(\hat{\beta}_0)$.
3. Encontrar una solución $\hat{\beta}_n$ de (2.17), para ρ acotada, usando un procedimiento iterativo comenzando con $\hat{\beta}_0$.

De este procedimiento, se obtiene un estimador que simultáneamente tiene alto punto de ruptura y alta eficiencia.

Daremos algunos detalles que deben tenerse en cuenta:

- El estimador inicial $\hat{\beta}_0$ debe ser de regresión, escala y transformaciones afines, lo que asegura que estas propiedades se trasladen al MM -estimador.
- El estimador de escala $\hat{\sigma}_n$ debe ser un M -estimador que satisfaga

$$\sum_{i=1}^n \rho_0 \left(\frac{r_i(\beta)}{\hat{\sigma}} \right) = 0.5.$$

Esto garantiza que el punto de ruptura de $\hat{\sigma}_n$ es 0.5.

- Debemos usar dos funciones ρ y ρ_0 diferentes, y cada una debe ser una ρ -función acotada.

Recordemos que los mínimos locales de $L(b)$ satisfacen (2.14). Para ρ satisfaciendo $\rho_0 \geq \rho$, Yohai (1987) demostró que si $\hat{\beta}_n$ es tal que

$$L(\hat{\beta}_n) \leq L(\hat{\beta}_0),$$

entonces $\hat{\beta}_n$ es consistente. También se puede mostrar que el punto de ruptura no es menor que el de $\hat{\beta}_0$. Además, si $\hat{\beta}_n$ es una solución de (2.14), entonces tiene la misma eficiencia que el mínimo global. Por lo tanto, no es necesario encontrar el mínimo absoluto de (2.17) para garantizar un alto punto de ruptura y alta eficiencia.

2.3.2. S -estimadores

Sea $\hat{\sigma}_n = \hat{\sigma}_n(\mathbf{r}(\mathbf{b}))$ un estimador robusto de escala equivariante basado en los residuos

$$\mathbf{r}(\mathbf{b}) = (r_1(\mathbf{b}), \dots, r_n(\mathbf{b})).$$

Luego, un estimador de regresión, puede ser definido como

$$\hat{\beta}_n = \arg \min_{\mathbf{b}} \hat{\sigma}_n(\mathbf{r}(\mathbf{b})). \quad (2.18)$$

Estos estimadores resultan equivariantes por regresión, traslaciones y cambios de escala.

Un caso particular de (2.18) es cuando $\hat{\sigma}(\mathbf{r})$ es un M -estimador de escala definido para cada \mathbf{r} por la solución de

$$\sum_{i=1}^n \rho_0 \left(\frac{r_i}{\hat{\sigma}_n} \right) = \delta,$$

donde ρ_0 es una ρ -función acotada. El estimador que resulta en este caso de (2.18) es llamado S -estimador (Rousseeauw y Yohai, 1984).

Cuando computamos un estimador inicial $\hat{\beta}_0$ para un MM -estimador, es muy recomendable usar una ρ perteneciente a la familia bicuadrada.

Desafortunadamente, un S -estimador con una función ρ suave, no tiene simultáneamente alto punto de ruptura y alta eficiencia, pero es de gran utilidad para ser utilizado como estimador inicial necesario para obtener un MM -estimador.

Capítulo 3

Estimación en el Modelo de Regresión Log–Gamma

Sea $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ una muestra aleatoria que cumple el modelo LGR, es decir cumple el modelo (1.2) con errores u_i con densidad de la forma (1.3) y moda en 0. Como vimos en el Capítulo 1, el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ lo podemos obtener minimizando la función *deviance* introducida en (1.5). De hecho

$$\arg \max_{\mathbf{b}} \prod_{i=1}^n g(y_i - \mathbf{x}'_i \mathbf{b}, \alpha) = \arg \max_{\mathbf{b}} \sum_{k=1}^n [\log(g(y_i - \mathbf{x}'_i \mathbf{b}, \alpha))]$$

o equivalentemente,

$$\arg \min_{\mathbf{b}} \sum_{i=1}^n [-\log(g(y_i - \mathbf{x}'_i \mathbf{b}, \alpha))] = \arg \min_{\mathbf{b}} \sum_{i=1}^n [\log g(0, \alpha) - \log(g(y_i - \mathbf{x}'_i \mathbf{b}, \alpha))],$$

con lo cual

$$\widehat{\boldsymbol{\beta}}_n^{ML} = \arg \min_{\mathbf{b}} \sum_{i=1}^n d^*(y_i, \mathbf{x}_i, \mathbf{b}),$$

donde d^* definida en (1.6) no depende del parámetro de forma α , pero sí depende su distribución.

Sea (\mathbf{x}', y) un vector aleatorio que tiene la misma distribución que los vectores de la muestra aleatoria, es decir cumple el modelo (1.2) con errores u cuya densidad es de la forma (1.3). Luego, si $E(\mathbf{xx}')$ es no singular

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}) \xrightarrow{D} N\left(0, \frac{1}{\alpha_0} E(\mathbf{xx}')^{-1}\right). \quad (3.1)$$

Sin embargo, como se mencionó en la Introducción, en este modelo los estimadores de máxima verosimilitud son muy sensibles a la presencia de datos atípicos.

Si intentáramos aplicar directamente algunas de los estimadores robustos propuestos para el modelo de regresión lineal, encontraríamos la desventaja de que, al ser la distribución log–Gamma asimétrica, si bien los estimadores de la pendiente son consistentes, los estimadores de la ordenada al origen no lo son. En consecuencia, también pueden resultar no consistentes las estimaciones de algunas funciones paramétricas, tales como las predicciones de la media condicional.

Por esta razón, es necesario desarrollar estimadores *ad hoc* para el modelo LGR. Bianco, García Ben y Yohai (2005) extienden los *MM*-estimadores al caso en que los errores tienen una distribución en la clase de las familias exponenciales del tipo (1.4), que incluye la distribución log–Gamma. La definición de estos estimadores es semejante a la de los estimadores originales introducidos por Yohai (1987) con la diferencia de que los residuos ordinarios son reemplazados por los residuos basados en la deviance. Los *MM*-estimadores que proponen dichos autores resultan Fisher-consistentes y poseen a la vez alto punto de ruptura y alta eficiencia.

A continuación se presenta la generalización de los *MM*-estimadores y algunas de sus propiedades más relevantes. La demostración de los resultados enunciados puede verse en Bianco, García Ben y Yohai (2005).

3.1. *M*-estimadores

Definimos un *M*-estimador de $\boldsymbol{\beta}$ como

$$\widehat{\boldsymbol{\beta}}_n^M = \arg \min_{\mathbf{b}} \sum_{i=1}^n \rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{\widehat{c}_n} \right), \quad (3.2)$$

donde \widehat{c}_n es un estimador de una constante de calibración c_0 (“tuning” constante) que depende de la muestra (\mathbf{x}'_i, y_i) , $1 \leq i \leq n$.

Asumimos que la función de pérdida ρ satisface la siguiente condición

A1.

- i) $\rho(0) = 0$.
- ii) Si $\alpha = \sup \rho(u)$, entonces $0 < \alpha < \infty$.
- iii) Si $0 \leq u < v$, $\rho(u) \leq \rho(v)$.
- iv) ρ es diferenciable.
- v) Si $\rho(u) < \alpha$ y $0 \leq u < v$, luego $\rho(u) < \rho(v)$.

En el caso del modelo (1.2) el M -estimador definido en (3.2) es Fisher-consistente, es decir, para cualquier α , $\boldsymbol{\beta}$ y c_0

$$\boldsymbol{\beta} = \arg \min_{\mathbf{b}} E_{\alpha, \beta} \left[\rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{c_0} \right) \right].$$

Esta propiedad fue probada para el caso de errores con función de densidad f_0 que satisfacen la siguiente condición

B. La densidad f_0 es estrictamente unimodal, continua y $f_0(t) > 0 \forall t$.

Es fácil verificar que la función de densidad de una variable aleatoria con distribución $\log \Gamma(\alpha, 1)$ satisface la propiedad anterior.

Otra forma de definir los M -estimadores es a través de la ecuación en derivadas. Sean

$$\psi(t) = \rho'(t), \quad w(t) = \frac{\psi(t)}{t},$$

y definamos los residuos $r(y_i, \mathbf{x}_i, \boldsymbol{\beta}) = y_i - \mathbf{x}_i' \boldsymbol{\beta}$. Luego, derivando en (3.2), resulta que el M -estimador $\hat{\boldsymbol{\beta}}_n^M$ satisface la siguiente ecuación

$$\sum_{i=1}^n \psi \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \hat{\boldsymbol{\beta}}_n^M)}}{\hat{c}_n} \right) \frac{(1 - \exp(r(y_i, \mathbf{x}_i, \hat{\boldsymbol{\beta}}_n^M)))}{\sqrt{d^*(y_i, \mathbf{x}_i, \hat{\boldsymbol{\beta}}_n^M)}} \mathbf{x}_i = 0$$

o equivalentemente

$$\sum_{i=1}^n w \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\beta}}_n^M)}}{\widehat{c}_n} \right) (1 - \exp(r(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\beta}}_n^M))) \mathbf{x}_i = 0. \quad (3.3)$$

Para obtener la distribución asintótica de $\widehat{\boldsymbol{\beta}}_n^M$, es necesario suponer las siguientes dos condiciones adicionales

A2. ρ es dos veces continuamente diferenciable y existe m tal que si $|u| \geq m$ entonces $\rho(u) = \sup \rho$.

A3. $E(\|\mathbf{x}\|^2) < \infty$, donde $\|\cdot\|$ es la norma l_2 y $E(\mathbf{xx}')$ es no singular.

La normalidad asintótica del M -estimador está dada por el siguiente teorema.

Teorema 3.1 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores aleatorios que satisfacen el modelo LGR con parámetros α y $\boldsymbol{\beta}$, y asumamos que se cumplen las condiciones **A1**- **A3**. Sea $\widehat{\boldsymbol{\beta}}_n$ una sucesión de estimadores que satisface (3.3) y, además, $\widehat{\boldsymbol{\beta}}_n \xrightarrow{c.s.} \boldsymbol{\beta}$. Si $\widehat{c}_n \xrightarrow{p} c_0$, luego

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} N \left(0, \frac{B(\psi, \alpha, c_0)}{(A(\psi, \alpha, c_0))^2} E(\mathbf{xx}')^{-1} \right),$$

donde

$$A(\psi, a, c) = E_a \left[w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} + w \left(\frac{\sqrt{h(u)}}{c} \right) e^u \right] \quad (3.4)$$

$$B(\psi, a, c) = E_a \left[w^2 \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u)^2 \right], \quad (3.5)$$

con u variable aleatoria $\log \Gamma(a, 1)$ y

$$h(u) = -u + e^u - 1. \quad (3.6)$$

Observación 3.1 : $E(e^u) = 1$.

Notemos que el estimador de máxima verosimilitud $\widehat{\beta}_n^{ML}$ es un M -estimador con función de pérdida $\rho_{MV} = u^2$. Es inmediato verificar que

$$\frac{B(\psi_{MV}, \alpha, c_0)}{(A(\psi_{MV}, \alpha, c_0))^2} = \frac{1}{\alpha}.$$

De este resultado se desprende que la eficiencia asintótica del M -estimador, para el modelo LGR, respecto del estimador de máxima verosimilitud está dada por

$$ARE(\psi, \alpha, c_0) = \frac{1}{\alpha} \frac{(A(\psi_M, \alpha, c_0))^2}{B(\psi_M, \alpha, c_0)}.$$

Supongamos que además de verificarse **A1-A3**, se cumple la siguiente condición adicional

$$\mathbf{A4.} \quad \psi(0) = 0, \quad \psi'(0) > 0, \quad \psi''(0) = 0,$$

que implica que $\rho(u)$ se comporta aproximadamente como u^2 en un entorno del 0. En ese caso, es posible demostrar que

$$\lim_{c \rightarrow \infty} ARE(\psi, \alpha, c) = 1.$$

Notemos que como generalmente $ARE(\psi, \alpha, c)$ es creciente y continua como función de c , podemos elegir $c = C_{ef}(\alpha)$ para obtener la eficiencia deseada. El valor $C_{ef}(a)$ se define como

$$ARE(\psi, a, C_{ef}(a)) = ef.$$

Observemos que la elección de un valor c para que el M -estimador tenga eficiencia ef , implica estimar α . El MM -estimador que definiremos más adelante incluye un estimador para α y por lo tanto $C_{ef}(\alpha)$ podrá ser estimado y así el MM -estimador será calibrado para obtener la eficiencia deseada.

3.2. S -estimadores

Como hemos visto en la sección anterior, para calibrar el M -estimador, es necesario un estimador inicial del parámetro de forma α . Para esto computaremos un estimador inicial de β , por medio de un S -estimador.

Para cada valor \mathbf{b} , sea $s_n(\mathbf{b})$ el M -estimador de escala de $\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}$ dado por

$$\sum_{i=1}^n \rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{s_n(\mathbf{b})} \right) = \delta,$$

donde ρ satisface **A1**. Para obtener un estimador con alto punto de ruptura, tomamos $\delta = \frac{1}{2} \sup \rho$.

El S -estimador de β para el modelo LGR, se define por

$$\tilde{\beta}_n = \arg \min_{\mathbf{b}} s_n(\mathbf{b}) \quad (3.7)$$

y el correspondiente estimador de escala

$$\hat{s}_n = \min_{\mathbf{b}} s_n(\mathbf{b}). \quad (3.8)$$

Definimos el valor asintótico de $s_n(\mathbf{b})$, como el valor $s(\alpha, \mathbf{b}, \beta)$ tal que

$$E_{\alpha, \beta} \left[\rho \left(\frac{\sqrt{d^*(y, \mathbf{x}, \mathbf{b})}}{s(\alpha, \mathbf{b}, \beta)} \right) \right] = \delta$$

Es fácil verificar que $s(\alpha, \mathbf{b}, \beta) = s(\alpha, \mathbf{b} - \beta, 0) = s^*(\alpha, \mathbf{b} - \beta)$.

Los siguientes teoremas prueban la consistencia de Fisher y consistencia fuerte de los S -estimadores.

Teorema 3.2 *Sea (\mathbf{x}', y) un vector aleatorio que satisface el modelo LGR con parámetros α y β y asumamos que ρ satisface **A1**. Luego, el S -estimador es Fisher-consistente, es decir*

$$\beta = \arg \min s^*(\alpha, \mathbf{b} - \beta).$$

Llamemos $S^*(a) = s^*(a, \mathbf{0})$, luego

$$E_a \left[\rho \left(\frac{\sqrt{h(u)}}{S^*(a)} \right) \right] = \delta, \quad (3.9)$$

donde $h(u)$ fue definida en (3.6) y u sigue una distribución $\log \Gamma(a, 1)$.

Para obtener el próximo resultado será necesario suponer una hipótesis adicional

$$\mathbf{A5}. P(\mathbf{x}'\theta = 0) < 1/2 \quad \forall \theta \in \mathbb{R}^p, \theta \neq 0.$$

Teorema 3.3 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores aleatorios independientes que satisfacen el modelo LGR con parámetros α y β y asumamos que se satisfacen las propiedades **A1-A3** y **A5**. Luego,

- (a) el S -estimador $\tilde{\beta}_n$ definido en (3.7) es fuertemente consistente para β .
- (b) el S -estimador de escala \hat{s}_n definido en (3.8) es fuertemente consistente para $S^*(\alpha)$.

Observemos que el S -estimador definido en (3.7) cumple

$$\tilde{\beta}_n = \arg \min_{\mathbf{b}} \sum_{I=1}^n \rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{s_n(\mathbf{b})} \right)$$

y por esta razón también es un M -estimador, de aquí que su distribución asintótica esté dada por el Teorema 3.1, con $c_0 = S^*(\alpha)$.

Usaremos la parte (b) del Teorema 3.3 para definir un estimador robusto para α . Para ello, necesitaremos el siguiente resultado.

Teorema 3.4 Supongamos que la función ρ satisface **A1**, entonces la función $S^*(a) : (0, \infty) \rightarrow (0, \infty)$ definida en (3.9) verifica

- (i) S^* es continua y estrictamente decreciente.
- (ii) $\lim_{a \rightarrow 0} S^*(a) = \infty$
- (iii) $\lim_{a \rightarrow \infty} S^*(a) = 0$.

El Teorema 3.4 implica que $S^*(a)$ es suryectiva, y por lo tanto, podemos computar la inversa de S^* , la cual llamaremos S^{*-1} .

Como vimos anteriormente \widehat{s}_n es un estimador consistente para $S^*(\alpha)$, con lo cual, bajo las condiciones del Teorema 3.3, un estimador consistente para α será

$$\widehat{\alpha}_n = S^{*-1}(\widehat{s}_n).$$

Finalmente, una medida de la robustez de un estimador es el punto de ruptura. Podemos definir de manera informal al punto de ruptura, como la menor proporción de outliers que llevan al estimador a romperse, es decir que puede tomar un valor más allá de cualquier límite y en ese sentido deja de ser informativo.

Trabajaremos con la versión de punto de ruptura para muestras finitas introducida por Donoho y Huber (1983).

Sea Z una muestra de n elementos, $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$, $\mathbf{z}_i = (\mathbf{x}'_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$.

Sea Z_m el conjunto de todas las muestras $Z^* = \{\mathbf{z}_1^*, \dots, \mathbf{z}_n^*\}$, tal que $\#\{i : \mathbf{z}_i^* = \mathbf{z}_i\} = n - m$, es decir, el conjunto de todas las muestras con $n - m$ elementos en común con Z .

Dado un estimador $\widehat{\beta}_n$ del parámetro de regresión, sean

$$b_m(Z, \widehat{\beta}_n) = \sup\{\|\widehat{\beta}_n(Z^*)\|, Z^* \in Z_m\}$$

y

$$m^* = \inf\{m : b_m(Z, \widehat{\beta}_n) = \infty\}.$$

Luego, el punto de ruptura finito (FSBDP) de $\widehat{\beta}_n$ en Z es $\epsilon^*(Z, \widehat{\beta}_n) = m^*/n$.

En nuestro caso, el estimador de escala \widehat{s}_n toma valores entre cero e infinito, por este motivo podemos definir dos FSBDP, uno a cero y otro a infinito de la siguiente forma

$$b_m^+(Z, \widehat{s}_n) = \sup\{\widehat{s}_n(Z^*), Z^* \in Z_m\}, b_m^-(Z, \widehat{s}_n) = \inf\{\widehat{s}_n(Z^*), Z^* \in Z_m\},$$

y

$$m^* = \min\{m : b_m^+(Z, \widehat{s}_n) = \infty\}, m^{**} = \min\{m : b_m^-(Z, \widehat{s}_n) = 0\}.$$

Luego, el FSBDP de \widehat{s}_n a infinito es $\epsilon^*(Z, \widehat{s}_n) = m^*/n$ y el FSBDP a cero es $\epsilon^{**}(Z, \widehat{s}_n) = m^{**}/n$.

Observemos que el parámetro α también toma valores entre cero e infinito, por lo cual podemos definir para $\widehat{\alpha}_n$ dos FSBDP, al igual que hicimos con \widehat{s}_n .

Diremos que el conjunto Z está en *posición general* si las covariables correspondientes a cualquier subconjunto de p observaciones son linealmente independientes. Bianco, Garcia Ben y Yohai (2005) probaron que si Z está en posición general y $\delta = \frac{1}{2} \sup \rho$, para valores de n grandes, los FSBDP de $\widetilde{\beta}_n$ y \widehat{s}_n son cercanos a 0.5.

3.3. MM-estimadores

Los *MM*-estimadores para el modelo LGR son una extensión natural de los *MM*-estimadores introducidos por Yohai (1987) para regresión ordinaria.

Supongamos que se quiere un estimador con eficiencia ef . Luego, el *MM*-estimador se define en los siguientes tres pasos:

- **Paso 1.** Computamos un S -estimador $\widetilde{\beta}_n$ y su correspondiente estimador de escala \widehat{s}_n , tomado $\delta = \frac{1}{2} \sup \rho$.
- **Paso 2.** Computamos $\widehat{\alpha}_n = S^{*-1}(\widehat{s}_n)$ y

$$\widehat{c}_n = \text{máx}(\widehat{s}_n, C_{ef}(\widehat{\alpha}_n)) = \text{máx}(\widehat{s}_n, C_{ef}(S^{*-1}(\widehat{s}_n)))$$

- **Paso 3.** El *MM*-estimador se define como un mínimo local $\widehat{\beta}_n$ de $R_n(\mathbf{b})$, donde

$$R_n(\mathbf{b}) = \sum_{i=1}^n \rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{c_n(\mathbf{b})} \right),$$

satisfaciendo $R_n(\widehat{\beta}_n) \leq R_n(\widetilde{\beta}_n)$.

Observación 3.2 : Cuando ρ pertenece a la familia de funciones bicuadradas dadas en (2.6), se puede mostrar numéricamente que $C_{ef}(S^{*-1}(s)) > s$ y en ese caso,

$$\hat{c}_n = \max(\hat{s}_n, C_{ef}(S^{*-1}(\hat{s}_n))) = C_{ef}(S^{*-1}(\hat{s}_n)).$$

En la práctica, en el primer paso computamos un S -estimador aproximado mediante submuestras de conjuntos elementales, propuesto por Rousseeuw y Leroy (1987). En el tercer paso, computamos $\hat{\beta}_n$ usando el algoritmo iterativo de mínimos cuadrados ponderados, comenzando por el estimador $\tilde{\beta}_n$ calculado en el Paso 1. La descripción de estos dos algoritmos la daremos en la siguiente sección.

Como vimos en la sección anterior, la elección de $\delta = \frac{1}{2} \sup \rho$ en el primer paso garantiza que $\tilde{\beta}_n$ y $\hat{\alpha}_n$ tengan FSBDP cercano a 0.5 para cualquier muestra en posición general. Bianco, Garcia Ben y Yohai (2005) demostraron que la cota inferior del punto de ruptura finito hallada para el S -estimador también lo es para el MM -estimador.

El siguiente Teorema nos da la distribución asintótica del MM -estimador.

Teorema 3.5 Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores aleatorios independientes que satisfacen el modelo LGR con parámetros α y β y asumamos que se satisfacen las propiedades **A1-A4**. Luego, la sucesión $\hat{\beta}_n$ de MM -estimadores satisface

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} N \left(0, \frac{B(\psi, \alpha, c_0)}{(A(\psi, \alpha, c_0))^2} E(\mathbf{xx}')^{-1} \right)$$

donde $A(\psi, \alpha, c)$ y $B(\psi, \alpha, c)$ están dados por (3.4) y (3.5) respectivamente y $c_0 = \max\{s^*(\alpha), C_{ef}(\alpha)\}$.

De acuerdo con el resultado anterior obtenemos que la eficiencia asintótica de $\hat{\beta}_n$ está dada por

$$\text{ARE}(\psi, \alpha, c_0) \geq ef .$$

Por lo tanto, el MM -estimador $\widehat{\beta}_n$ tiene simultáneamente punto de ruptura finito cercano a 0.5 para muestras en posición general y eficiencia asintótica mayor o igual a ef .

3.4. Aspectos computacionales

Supongamos que queremos un MM -estimador con eficiencia ef y punto de ruptura 0.5 para el parámetro β del modelo LGR. Empezamos calculando la función $S^*(a)$ y $C_{ef}(a)$ numéricamente. Para esto, primero computamos estas funciones sobre una grilla y luego calculamos suavizados basado en *splines* para computarlas en todo punto, como estas funciones no dependen de la muestra, serán calculadas una única vez.

El procedimiento para computar un MM -estimador para el modelo LGR es similar al usado para regresión ordinaria (Yohai, 1987).

El estimador inicial $\widetilde{\beta}_n$ es un S -estimador aproximado basado en N submuestras aleatorias de tamaño p , este método será explicado en detalle más abajo.

Una vez obtenido el estimador $\widetilde{\beta}_n$ y su correspondiente \widehat{s}_n , calculamos $\widehat{\alpha}_n = S^{*-1}(\widehat{s}_n)$ y $\widehat{c}_n = \max(\widehat{s}_n, C_e(\widehat{\alpha}_n))$.

Para obtener el M -estimador final, usaremos el método de mínimos cuadrados pesados iterados, que describimos a continuación.

De acuerdo a la ecuación (3.3) el M -estimador $\widehat{\beta}_n^M$ satisface

$$\sum_{i=1}^n w \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)}}{\widehat{c}_n} \right) (1 - \exp(r(y_i, \mathbf{x}_i, \widehat{\beta}_n^M))) \mathbf{x}_i = 0 ,$$

multiplicando y dividiendo cada término por $r(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)$, obtenemos que esta ecuación es equivalente a

$$\sum_{i=1}^n w_i^* (y_i - \mathbf{x}_i' \widehat{\beta}_n^M) \mathbf{x}_i = 0,$$

donde los pesos w_i^* están dados por

$$w_i^* = w \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\beta}}_n^M)}}{\widehat{c}_n} \right) \frac{(1 - \exp(r(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\beta}}_n^M)))}{r(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\beta}}_n^M)}$$

El procedimiento de IRWLS, que depende de un parámetro de tolerancia ϵ , es el siguiente

1. Computar un estimador inicial $\widehat{\boldsymbol{\beta}}_0$ de $\boldsymbol{\beta}$ (en nuestro caso será $\widetilde{\boldsymbol{\beta}}_n$) y un estimador de c , \widehat{c}_0 .
2. Para $k = 0, 1, 2, \dots$:
 - (a) Dados $\widehat{\boldsymbol{\beta}}_k$, computar $r_{i,k} = y_i - \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_k$ y $w_{i,k}^*$
 - (b) Computar $\widehat{\boldsymbol{\beta}}_{k+1}$ resolviendo

$$\sum_{i=1}^n w_{i,k}^* (y_i - \mathbf{x}_i' \mathbf{b}) \mathbf{x}_i = 0.$$

3. Parar cuando $\|\widehat{\boldsymbol{\beta}}_{k+1} - \widehat{\boldsymbol{\beta}}_k\| / \|\widehat{\boldsymbol{\beta}}_k\| < \epsilon$

Respecto al método de submuestreo recordemos que para encontrar un S -estimador, debemos resolver

$$\widetilde{\boldsymbol{\beta}}_n = \arg \min_{\mathbf{b}} s_n(\mathbf{b}),$$

con $s_n(\mathbf{b})$ tal que

$$\sum_{i=1}^n \rho \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{s_n(\mathbf{b})} \right) = \delta.$$

Para encontrar un mínimo aproximado a esta ecuación, computaremos una gran cantidad N de conjuntos candidatos y reemplazaremos la minimización sobre todos los $\mathbf{b} \in \mathbb{R}^p$ por la minimización sobre este gran conjunto finito.

Para computar las soluciones posibles, tomamos submuestras de tamaño p de la forma:

$$\{(\mathbf{x}'_i, y_i) : i \in J\}, \quad J \subset \{1, \dots, n\}, \quad \#(J) = p.$$

Para cada J obtenemos el valor de \mathbf{b} que ajusta los p puntos. Luego, el problema de minimizar sobre todos los $\mathbf{b} \in \mathbb{R}^p$, es reemplazado por el problema de minimizar sobre J .

Elegir todas las $\binom{n}{p}$ submuestras, puede ser demasiado grande, es por esto que tomamos N de ellas al azar: $\{J_k : k = 1, \dots, N\}$ y obtenemos los estimadores $\tilde{\beta}_1, \dots, \tilde{\beta}_N$ y sus correspondientes estimadores de escala $\hat{s}_1, \dots, \hat{s}_N$. Luego, el S -estimador aproximado es $\tilde{\beta}_n = \beta_{i_0}$, donde $s_{i_0} = \min(s_i, 1 \leq i \leq N)$.

La elección del valor de N es similar al caso de regresión ordinaria y es tratado en Rousseeuw y Leroy (1987).

Capítulo 4

Test de Tipo Wald en el Modelo de Regresión Log–Gamma

4.1. Estadístico de Wald

Una de las herramientas más utilizadas en inferencia estadística es la de test de hipótesis, que es un método de decisión que permite elegir de manera formal entre dos opciones que conciernen a la distribución de una población.

Enfocaremos el problema de testear bajo el modelo LGR definido en (1.2), hipótesis de tipo

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0.$$

Como en regresión lineal, es frecuente que las hipótesis involucren sólo a q de los p parámetros. Sea $\boldsymbol{\beta}' = (\boldsymbol{\beta}'_{(1)}, \boldsymbol{\beta}'_{(2)})$, $\widehat{\boldsymbol{\beta}}' = (\widehat{\boldsymbol{\beta}}'_{(1)}, \widehat{\boldsymbol{\beta}}'_{(2)})$ y $\mathbf{x}' = (\mathbf{x}'_{(1)}, \mathbf{x}'_{(2)})$, donde $\boldsymbol{\beta}_{(1)} \in \mathbb{R}^q$ y $\boldsymbol{\beta}_{(2)} \in \mathbb{R}^{p-q}$. Los resultados que presentaremos podrán extenderse a situaciones en las que interesa testear hipótesis de la forma

$$H_0^* : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0} \quad \boldsymbol{\beta}_{(2)} \text{ sin especificar.}$$

Una forma natural de testear H_0 es por medio de un estadístico de tipo Wald, que mide la discrepancia entre el estimador $\widehat{\boldsymbol{\beta}}$ y el valor del parámetro bajo la hipótesis nula, usando la forma cuadrática

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{4.1}$$

donde $\widehat{\Sigma}$ es un estimador consistente de la matriz de covarianza de $\widehat{\beta}$.

Como vimos en el capítulo anterior, una forma de estimar β es por medio del estimador clásico basado en el método de máxima verosimilitud. Sin embargo, los procedimientos de inferencia basados en este método heredan su sensibilidad a datos atípicos; en efecto, una pequeña proporción de estos puede afectar severamente tanto el nivel como la potencia de los tests derivados. Por esta razón, es necesario buscar métodos más resistentes. El propósito en este sentido es doble ya que es deseable obtener tests de hipótesis que bajo contaminación conserven tanto un nivel estable como una alta potencia. Como fue mencionado en la Introducción, los trabajos de Heritier y Ronchetti (1994) y Cantoni y Ronchetti (2001) investigan en esta dirección. Heritier y Ronchetti (1994) proponen tests robustos para un modelo paramétrico general. Por otro lado, Cantoni y Ronchetti (2001), usando la noción de cuasi-verosimilitud, definen versiones robustas de la deviance e introducen una familia de tests para el problema de selección de variables en modelos lineales generalizados.

En este capítulo se propone estudiar algunas propiedades asintóticas de tests de tipo Wald basados en los MM -estimadores descritos en el Capítulo 3 y se muestra que bajo la hipótesis nula tienen el mismo comportamiento asintótico que su equivalente clásico, basado en los estimadores de máxima verosimilitud.

4.2. Comportamiento asintótico

Para poder establecer la región crítica de un test del tipo (4.1), necesitamos las distribuciones muestrales de los estadísticos involucrados.

En esta sección estudiaremos el comportamiento asintótico del estadístico de un test del tipo (4.1) basado en los estimadores de los parámetros, tanto para el caso clásico como el robusto.

4.2.1. Estadístico clásico

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores independientes que satisfacen el modelo LGR. Como vimos en el Capítulo 3, cuando $E(\mathbf{xx}')$ es definida positiva, el estimador de máxima verosimilitud $\widehat{\boldsymbol{\beta}}_n^{ML}$ tiene distribución asintótica dada por (3.1), donde la matriz de covarianza asintótica es de la forma

$$\Sigma = \frac{1}{\alpha} E(\mathbf{xx}')^{-1}.$$

El siguiente teorema establece el comportamiento del estadístico de Wald basado en el estimador clásico de $\boldsymbol{\beta}$.

Teorema 4.1

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores independientes que satisfacen el modelo LGR. Sea $\widehat{\Sigma}_n$ una sucesión consistente de estimadores de Σ , la matriz de covarianza asintótica de $\widehat{\boldsymbol{\beta}}_n^{ML}$. Luego,

- i) Bajo $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0) \xrightarrow{D} \chi_p^2$
- ii) Bajo $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0) \xrightarrow{p} \infty$, para cualquier $\boldsymbol{\beta}$ fijo.

Demostración.

- i) La demostración es inmediata a partir de (3.1) y del hecho de que $\widehat{\Sigma}_n \xrightarrow{p} \Sigma$.

- ii) Sea

$$\begin{aligned} nD^{ML}(\boldsymbol{\beta}_0, \widehat{\Sigma}_n) &= n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0) \\ &= n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta} + \boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &= n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}) + n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &= n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta} + \boldsymbol{\beta} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}) + n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &= nD^{ML}(\boldsymbol{\beta}, \widehat{\Sigma}_n) + n(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}) + n(\widehat{\boldsymbol{\beta}}_n^{ML} - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \end{aligned}$$

como $\widehat{\Sigma}_n^{-1}$ es simétrica,

$$nD^{ML}(\beta_0, \widehat{\Sigma}_n) = nD^{ML}(\beta, \widehat{\Sigma}_n) + n(\widehat{\beta}_n^{ML} - \beta + \widehat{\beta}_n^{ML} - \beta_0)' \widehat{\Sigma}_n^{-1} (\beta - \beta_0).$$

Luego, por un lado, $nD^{ML}(\beta, \widehat{\Sigma}_n) \xrightarrow{D} \chi_p^2$ es un término positivo con probabilidad tendiendo a uno y por otro lado, usando que $\widehat{\Sigma}_n$ y $\widehat{\beta}_n^{ML}$ son estimadores consistentes, tenemos que

$$(\widehat{\beta}_n^{ML} - \beta + \widehat{\beta}_n^{ML} - \beta_0)' \widehat{\Sigma}_n^{-1} (\beta - \beta_0) \xrightarrow{p} (\beta - \beta_0)' \Sigma^{-1} (\beta - \beta_0) > 0,$$

con lo cual $nD^{ML}(\beta_0, \widehat{\Sigma}_n) \xrightarrow{p} \infty$. ■

Notemos que el resultado anterior requiere de un estimador consistente de la matriz de covarianza asintótica de $\widehat{\beta}_n^{ML}$, que involucra al parámetro de forma α , por lo tanto será necesaria su estimación.

Dado $\widehat{\alpha}$ un estimador consistente de α , podríamos estimar a Σ mediante

$$\widehat{\Sigma}_n = \frac{1}{\widehat{\alpha}} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right\}^{-1},$$

que por la Ley de los Grandes Números y la continuidad será consistente.

Una posible forma de estimar a α es mediante el método de los momentos, tal como describimos a continuación.

Dada $\mathbf{z} = (z_1, \dots, z_n)$ una colección de variables aleatorias independientes tales que $z_i \sim \Gamma(\alpha, \mu_i)$, definamos las variables $y_i = \frac{z_i}{\mu_i} - 1$. Es fácil ver que

$$E(y_i) = 0, \text{VAR}(y_i) = \frac{1}{\alpha} \text{ y } E(y_i^2) = \text{VAR}(y_i).$$

El estimador de momentos de α puede hallarse igualando el segundo momento de y_i con el segundo momento muestral, es decir planteando

$$\frac{1}{\alpha} = E(y_i^2) = E_\alpha \left[\left(\frac{z_i}{\mu_i} - 1 \right)^2 \right] = \frac{1}{n} \sum_{i=1}^n \left(\frac{z_i}{\mu_i} - 1 \right)^2,$$

por lo tanto

$$\frac{1}{\widehat{\alpha}_n} = \frac{1}{n} \sum_{i=1}^n \left(\frac{z_i}{\mu_i} - 1 \right)^2.$$

Es sencillo verificar, por la Ley de los Grandes Números, que $\widehat{\alpha}_n$ resulta consistente y el mismo resultado es válido si μ_i es estimado en forma consistente. Esto último puede lograrse en el caso de regresión estimando a μ_i mediante $\widehat{\mu}_i = \exp\left(\mathbf{x}'_i \widehat{\boldsymbol{\beta}}_n^{ML}\right)$, y suele reemplazarse el factor $1/n$ por $1/n - p$, obteniendo

$$\frac{1}{\widehat{\alpha}_n} = \frac{1}{n - p} \sum_{i=1}^n \left(\frac{z_i}{\widehat{\mu}_i} - 1 \right)^2.$$

4.2.2. Estadístico robusto

Consideremos $\widehat{\boldsymbol{\beta}}_n$ una sucesión de *MM*-estimadores tal como fueron definidos por el algoritmo de tres pasos en la Sección 3.3. El Teorema 3.5 establece que la distribución asintótica de los *MM*-estimadores $\widehat{\boldsymbol{\beta}}_n$ es normal con matriz de covarianza

$$\frac{B}{A^2} E(\mathbf{xx}')^{-1},$$

donde A y B están dados por

$$A = E_a \left[w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} + w \left(\frac{\sqrt{h(u)}}{c} \right) e^u \right]$$

$$B = E_a \left[w^2 \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u)^2 \right],$$

y $c_0 = \max\{s^*(\alpha), C_{ef}(\alpha)\}$. Por lo tanto, el estadístico de Wald basado en $\widehat{\boldsymbol{\beta}}_n$ para testear

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad vs. \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0,$$

será

$$nD(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\Sigma}}_n) = n(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \widehat{\boldsymbol{\Sigma}}_n^{-1} (\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0).$$

Dados \widehat{A}_n y \widehat{B}_n , estimadores consistentes de A y B respectivamente, por la Ley de los Grandes Números, Σ puede ser estimado en forma consistente por medio de

$$\widehat{\Sigma}_n = \frac{\widehat{B}_n}{\widehat{A}_n^2} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right\}^{-1}.$$

Una forma natural de estimar a A y B es usando sus versiones muestrales, es decir reemplazando el valor esperado por el promedio muestral como sigue

$$\begin{aligned} \widehat{A}_n &= \frac{1}{n} \sum_{k=1}^n \left[w' \left(\frac{\sqrt{h(u_k)}}{\widehat{c}_n} \right) \frac{(1 - e^{u_k})^2}{2\widehat{c}_n \sqrt{h(u_k)}} + w \left(\frac{\sqrt{h(u_k)}}{\widehat{c}_n} \right) e^{u_k} \right] \\ \widehat{B}_n &= \frac{1}{n} \sum_{k=1}^n \left[w^2 \left(\frac{\sqrt{h(u_k)}}{\widehat{c}_n} \right) (1 - e^{u_k})^2 \right]. \end{aligned}$$

Nuestro primer objetivo será demostrar que los estimadores de A y B propuestos son consistentes. Para ello, usaremos el siguiente resultado que corresponde al Lema 4.2 de Yohai (1987).

Lema 4.2 (Yohai, 1987) *Sea $g : \mathbb{R}^k \times \mathbb{R}^h \rightarrow \mathbb{R}$ continua y sea Q una una función de distribución sobre \mathbb{R}^k tal que para algún $\delta > 0$ se tiene*

$$E \left[\sup_{|c-c_0| \leq \delta} |g(u, c)| \right] < \infty.$$

Sea $\widehat{\boldsymbol{\lambda}}_m$ una sucesión de estimadores en \mathbb{R}^h tal que $\lim_{m \rightarrow \infty} \widehat{\boldsymbol{\lambda}}_m = \boldsymbol{\lambda}_0$. Si z_1, z_2, \dots, z_n , son variables i.i.d en \mathbb{R}^k con distribución Q , tenemos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n g(z_i, \boldsymbol{\lambda}_n) = E_Q[g(z, \boldsymbol{\lambda}_0)].$$

Asumiremos además la siguiente hipótesis para demostrar la consistencia de los estimadores de A y B

A6. ψ satisface que $t\psi'(t) \leq \psi(t)$ y además $\eta(t) = \frac{w'(t)}{t}$ es acotada.

Observación 4.1 : *la primera condición dada en **A6** garantiza que $w(t)$ sea monótona decreciente y la segunda condición controla el decrecimiento de w' . Ambas condiciones son satisfechas, por ejemplo, por la función bicuadrada de Tukey.*

La siguiente proposición prueba la consistencia de los estimadores propuestos.

Proposición 4.1 *Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores independientes que satisfacen el modelo LGR. Si se cumplen **A1-A4** y **A6** y además $\widehat{c}_n \xrightarrow{p} c_0$, entonces*

$$(a) \quad \widehat{A}_n \xrightarrow{p} A$$

$$(b) \quad \widehat{B}_n \xrightarrow{p} B .$$

Demostración.

(a) Por el Lema 4.2 en Yohai (1987), basta ver que

$$E \left[\sup_{|c-c_0| \leq \delta} |g(u, c)| \right] < \infty$$

para

$$g(u, c) = w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} + w \left(\frac{\sqrt{h(u)}}{c} \right) e^u.$$

Para ver que la esperanza de la suma está acotada, veamos que la esperanza de cada sumando lo está.

Sea $c_0 - \delta < c < c_0 + \delta$, luego $\frac{1}{c^2} < \frac{1}{(c_0 - \delta)^2}$. Usando **A2** y **A6** tenemos que

$\exists m$ tal que $|\eta(t)| = |w'(t)/t| \leq \|\eta\|_\infty \mathbf{I}(|t| < m)$ y por lo tanto,

$$\begin{aligned} \sup_{|c-c_0| \leq \delta} \left| w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} \right| &\leq \|\eta\|_\infty \frac{(1 - e^u)^2}{2(c_0 - \delta)^2} \sup_{|c-c_0| \leq \delta} \left[\mathbf{I} \left(\left| \frac{\sqrt{h(u)}}{c} \right| < m \right) \right] \\ &\leq \|\eta\|_\infty \frac{(1 - e^u)^2}{2(c_0 - \delta)^2} \sup_{|c-c_0| \leq \delta} \left[\mathbf{I} \left(0 < \sqrt{h(u)} < mc \right) \right] \\ &\leq \|\eta\|_\infty \frac{(1 - e^u)^2}{2(c_0 - \delta)^2} \left[\mathbf{I} \left(0 < h(u) < (m(c_0 + \delta))^2 \right) \right] . \end{aligned}$$

Como $h(u) = -u + e^u - 1$, entonces $h'(u) = -1 + e^u = 0$ si y sólo si $u = 0$. Teniendo en cuenta que $h''(u) = e^u > 0$, resulta que h alcanza su valor mínimo, es decir 0, cuando $u = 0$ y además, se concluye que $h(u)$ es convexa. En consecuencia, existen $a < 0$ y $b > 0$ tales que

$$\sup_{|c-c_0| \leq \delta} \left| w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} \right| \leq \|\eta\|_\infty \frac{(1 - e^u)^2}{2(c_0 - \delta)^2} \mathbf{I}(a < u < b)$$

y tomando esperanza obtenemos que

$$E \left[\sup_{|c-c_0| \leq \delta} \left| w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} \right| \right] < \infty \quad \blacksquare$$

Veamos que el segundo sumando está acotado.

Por **A2** $w(t)$ está acotada, luego

$$\sup_{|c-c_0| \leq \delta} \left| w \left(\frac{\sqrt{h(u)}}{c} \right) e^u \right| \leq \|w\|_\infty e^u$$

tomando esperanza y usando la Observación 3.1, podemos concluir que

$$E \left[\sup_{|c-c_0| \leq \delta} \left| w \left(\frac{\sqrt{h(u)}}{c} \right) e^u \right| \right] \leq \|w\|_\infty \quad \blacksquare$$

(b) Sea ahora

$$g(u, c) = w \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u).$$

A6 implica que w es monótona decreciente, entonces tenemos

$$\sup_{|c-c_0| \leq \delta} \left| w \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u) \right| \leq \left| w \left(\frac{\sqrt{h(u)}}{(c_0 - \delta)} \right) (1 - e^u) \right|$$

Como por **A2**, $|w(t)| \leq \|w\|_\infty \mathbf{I}(|t| < m)$, resulta

$$\left| w \left(\frac{\sqrt{h(u)}}{(c_0 - \delta)} \right) (1 - e^u) \right| \leq \|w\|_\infty \mathbf{I} \left(0 < \frac{\sqrt{h(u)}}{(c_0 - \delta)} < m \right) (1 - e^u).$$

Nuevamente, razonando como en a), concluimos que existen $a' < 0$ y $b' > 0$ tales que

$$\sup_{|c-c_0| \leq \delta} \left| w \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u) \right| \leq \|w\|_\infty (1 - e^u) \mathbf{I}(a' < u < b'),$$

y tomando esperanza podemos concluir

$$E \left[\sup_{|c-c_0| \leq \delta} \left(w \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u) \right)^2 \right] < \infty \quad \blacksquare$$

En el siguiente Teorema mostramos la distribución asintótica del estadístico de Wald robusto bajo la hipótesis nula y una alternativa fija.

Teorema 4.2

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores independientes que satisfacen el modelo LGR. Si se cumplen **A1-A4** y **A6** y además, $\hat{\boldsymbol{\beta}}_n \xrightarrow{c.s.} \boldsymbol{\beta}$ y $\hat{\boldsymbol{\Sigma}}_n \xrightarrow{p} \boldsymbol{\Sigma}$, entonces

- i) Bajo $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \hat{\boldsymbol{\Sigma}}_n^{-1} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{D} \chi_p^2$
- ii) Bajo $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \hat{\boldsymbol{\Sigma}}_n^{-1} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{p} \infty$, para cualquier $\boldsymbol{\beta}$ fijo.

Demostración.

- i) La demostración es inmediata a partir del Teorema (3.5) y del hecho de que $\hat{\boldsymbol{\Sigma}}_n$ es un estimador consistente de $\boldsymbol{\Sigma}$.
- ii) Al igual que para el caso del estadístico clásico, tenemos que

$$\begin{aligned} nD(\boldsymbol{\beta}_0, \hat{\boldsymbol{\Sigma}}_n) &= n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \hat{\boldsymbol{\Sigma}}_n^{-1} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\ &= nD(\boldsymbol{\beta}, \hat{\boldsymbol{\Sigma}}_n) + n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} + \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \hat{\boldsymbol{\Sigma}}_n^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \end{aligned}$$

Por un lado, $nD(\boldsymbol{\beta}, \widehat{\Sigma}_n) \xrightarrow{D} \chi_p^2$, por lo tanto este término es positivo, con probabilidad tendiendo a uno. Además, usando la consistencia de $\widehat{\Sigma}_n$ y de $\widehat{\boldsymbol{\beta}}_n$ y el hecho de que Σ es una matriz definida positiva, obtenemos que

$$(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} + \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \widehat{\Sigma}_n^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \xrightarrow{p} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) > 0,$$

con lo cual $nD(\boldsymbol{\beta}_0, \widehat{\Sigma}_n) \xrightarrow{p} \infty$ ■

Como se mencionó anteriormente, en algunas situaciones estamos interesados en estudiar el impacto de q covariables sobre la variable de respuesta y . En ese caso, queremos hacer inferencia sólo sobre algunas de las componentes del parámetro de regresión $\boldsymbol{\beta}$. Es decir, nos interesará testear hipótesis de la forma

$$H_0^* : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0} \quad \boldsymbol{\beta}_{(2)} \text{ sin especificar.}$$

Si queremos testear H_0^* podemos usar el estadístico

$$nD_1(\boldsymbol{\beta}_{(1)}, \widehat{\Sigma}_{11}) = n(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})' \widehat{\Sigma}_{11}^{-1} (\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0}), \quad (4.2)$$

donde $\widehat{\Sigma}_{11}^{-1}$ es la submatriz de $\widehat{\Sigma}_n$ de tamaño $q \times q$, correspondiente a las coordenadas de $\widehat{\boldsymbol{\beta}}_{(1)}$.

De forma análoga al Teorema 4.2, se puede demostrar el siguiente teorema, que establece la distribución asintótica del estadístico del test propuesto.

Teorema 4.3

Sean $(\mathbf{x}'_1, y_1), \dots, (\mathbf{x}'_n, y_n)$ vectores independientes que satisfacen el modelo LGR, asumimos que se cumplen **A1-A4** y **A6**, $\widehat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$ y $\widehat{\Sigma}_n$ una sucesión consistente de estimadores de Σ . Luego,

- i) Bajo $H_0^* : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $nD_1(\boldsymbol{\beta}_{(1),0}, \widehat{\Sigma}_{11}) \xrightarrow{D} \chi_q^2$
- ii) Bajo $H_1^* : \boldsymbol{\beta}_{(1)} \neq \boldsymbol{\beta}_{(1),0}$, $nD_1(\boldsymbol{\beta}_{(1),0}, \widehat{\Sigma}_{11}) \xrightarrow{p} \infty$, para cualquier $\boldsymbol{\beta}_{(1)}$ fijo.

A partir de los Teoremas 4.2 y 4.3 concluimos que para testear las hipótesis nulas vistas con un nivel nominal τ , tenemos los siguientes test consistentes

- rechazo $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ si $nD(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\Sigma}}_n) > \chi_{p,\tau}^2$,

y

- rechazo $H_0^* : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$ si $nD_1(\boldsymbol{\beta}_{(1),0}, \widehat{\boldsymbol{\Sigma}}_{11}) > \chi_{q,\tau}^2$.

Capítulo 5

Simulaciones

Para estudiar el comportamiento del test propuesto y poder compararlo con el test clásico, realizamos un estudio de simulación, analizando estabilidad en el nivel y en la potencia frente a muestras contaminadas y sin contaminar.

5.1. Nivel

Para el estudio de nivel, generamos 5000 muestras independientes de tamaños $n = 200$, $n = 100$ y $n = 50$, con $p = 3$, es decir intercept y dos covariables.

Las covariables $\mathbf{x}'_i = (x_{1i}, x_{2i})$ fueron generadas con distribución normal multivariada, sin pérdida de generalidad fue elegida como $N_2(\mathbf{0}, \mathbf{I}_2)$ con $\mathbf{0}' = (0, 0)$ e \mathbf{I}_2 la matriz identidad en $\mathbb{R}^{2 \times 2}$, ya que el estimador es afín-equivariante. La variable de respuesta y_i satisface el modelo LGR

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 + u_i, \quad i = 1, \dots, n, \quad (5.1)$$

donde u_i tiene distribución $\log \Gamma(\alpha, 1)$, y $\boldsymbol{\beta}'_0 = (0, 0, 0)$. Consideramos como valores del parámetro de forma, $\alpha = 1, 3$ y 5 . Denotaremos C_0 al caso sin contaminación.

Bajo contaminación, trabajamos con muestras de tamaño $n = 100$ y contaminamos con distintas proporciones de outliers: 1%, 3% y 5%. Llamaremos C_1, C_3, C_5 a los casos contaminados con uno, tres y cinco outliers, respectivamente. En las muestras originales generadas según el modelo (5.1), introdujimos por reemplazo outliers de la forma (x_{10}, x_{20}, y_0) , cuyo efecto depende de la magnitud $x_{10}^2 + x_{20}^2$. Por esta razón, sin pérdida de generalidad, los puntos anómalos fueron tomados siguiendo el patrón $(x_0, 0, mx_0)$, donde m es la pendiente de contaminación de la primera variable independiente. Con el fin de generar puntos atípicos de diferente grado de severidad, elegimos distintos leverage: $x_0 = 1, 3$ y 5 y diferentes pendientes $m=0.5, 1, 1.5, 2, 2.5, 3$.

Para poder entender el efecto del parámetro de forma α sobre los outliers generados, mostramos las Figuras 5.1 y 5.2.

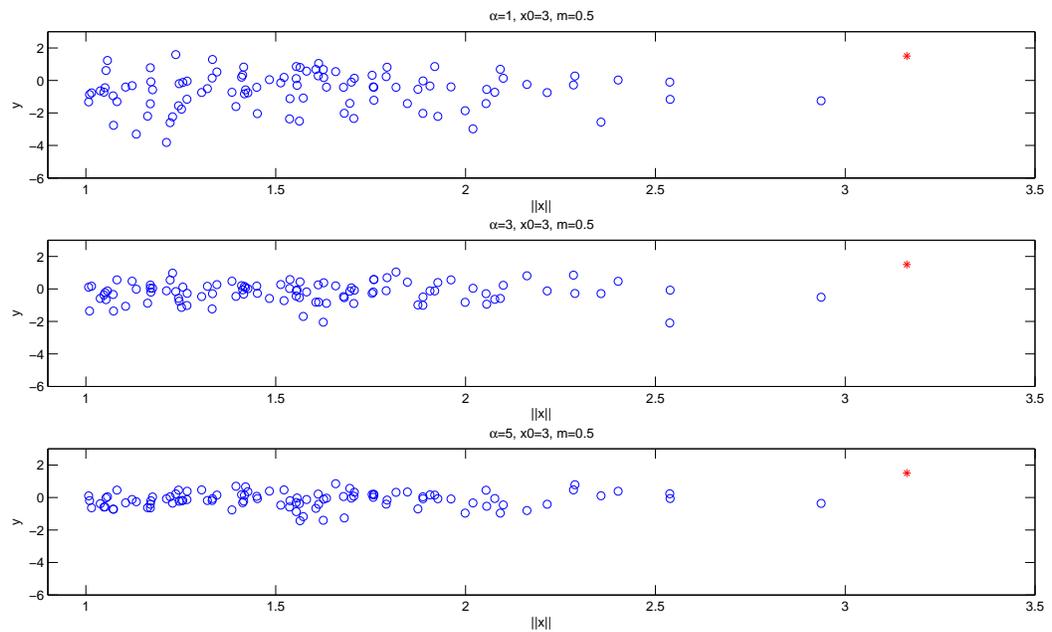


Figura 5.1: La muestra original se representa con 'o' y el outlier con '*'.

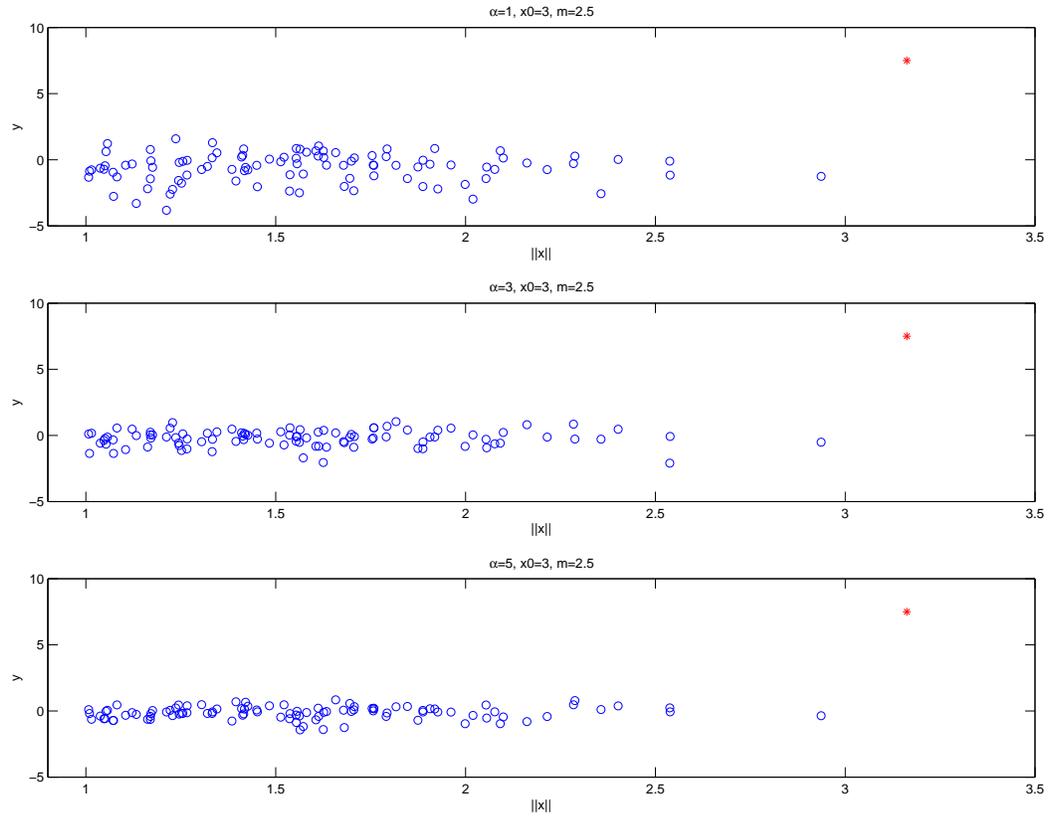


Figura 5.2: La muestra original se representa con 'o' y el outlier con '*'.

Observemos como para un mismo tipo de outliers el efecto que este produce sobre la muestra se incrementa al aumentar el valor de α , es decir, el mismo outlier aumenta su grado de severidad, cuando α toma valores más grandes.

En la definición del estimador dada en (3.3) utilizamos como función w

$$w(x) = \begin{cases} \frac{6}{k^2} \left(1 - \left(\frac{x}{k}\right)^2\right)^2 & \text{si } |x| \leq k \\ 0 & \text{si } |x| > k \end{cases} \quad (5.2)$$

basada en la función ρ bicuadrada de Tukey, con constante de calibración $k = 1$. Tomamos eficiencia asintótica para el MM -estimador $e = 0.9$.

El número de submuestras utilizado en el algoritmo para calcular el S -estimador inicial es 500.

La hipótesis nula considerada fue $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, con $\boldsymbol{\beta}'_0 = (0, 0, 0)$. Denotaremos D_R y D_{ML} a los estadísticos de los test de tipo Wald robusto y clásico, respectivamente.

El nivel de significación nominal es $\tau = 5\%$.

- Caso sin contaminar

En el Cuadro 5.1 se presentan las frecuencias de rechazo observadas bajo H_0 para distintos tamaños de muestra $n = 50, 100$ y 200 .

n	$\alpha = 1$		$\alpha = 3$		$\alpha = 5$	
	D_{ML}	D_R	D_{ML}	D_R	D_{ML}	D_R
50	0.0872	0.1202	0.0708	0.1148	0.0796	0.1238
100	0.0766	0.0752	0.0606	0.0774	0.0618	0.0804
200	0.0620	0.0648	0.0534	0.0586	0.0544	0.0626

Cuadro 5.1: Frecuencia observada de rechazo bajo H_0 en el caso C_0 , $\tau = 5\%$.

Al estudiar el nivel del test clásico y el propuesto, observamos que las frecuencias de rechazo en ambos casos son cercanas al valor nominal $\tau = 5\%$.

El test clásico da muy buenos resultados para cualquiera de los valores de n considerados, mientras que el test propuesto se acerca al nivel nominal a partir de $n = 100$.

- Caso contaminado

En las Figuras 5.3, 5.4 y 5.5 se muestran las frecuencias de rechazo obtenidas bajo H_0 variando la pendiente m de los outliers cuando se utilizan el test clásico y el robusto bajo los esquemas de contaminación C_1 , C_3 y C_5 , respectivamente, para los distintos valores de α y x_0 considerados. Con 'o' se representan las frecuencias

obtenidas con el test robusto y con '*' las obtenidas con el caso clásico; con la línea segmentada se representa el valor nominal $\tau = 0.05$.

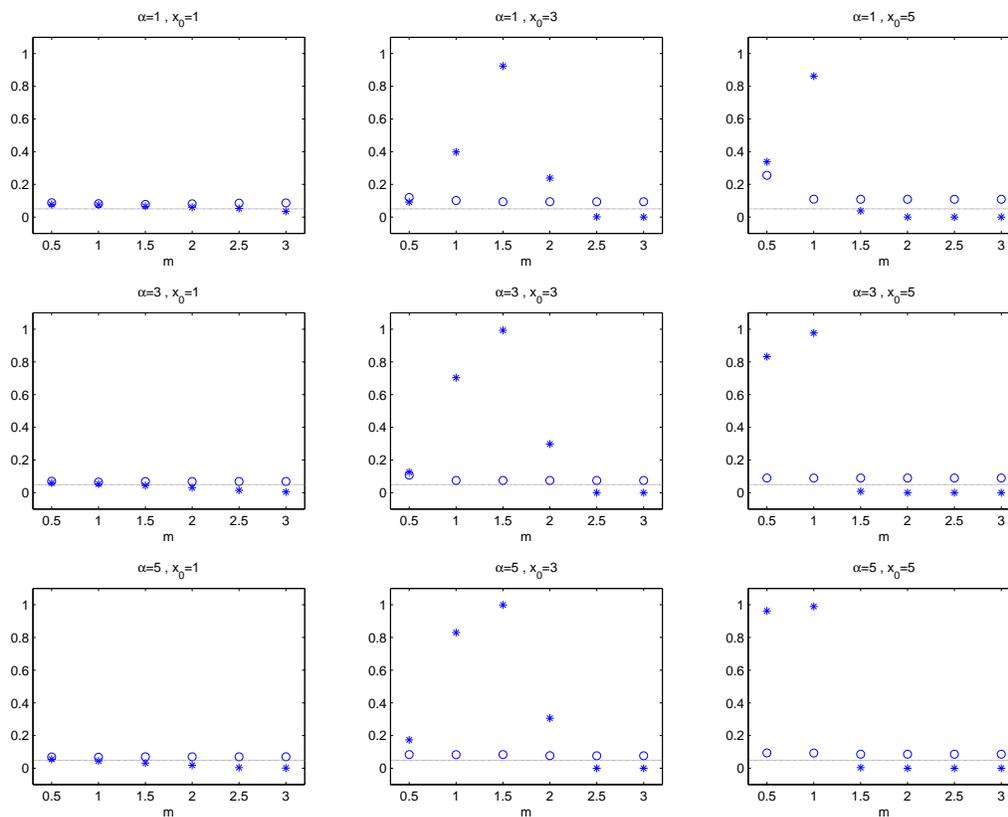


Figura 5.3: Frecuencia observada de rechazo bajo H_0 en el caso C_1 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

En la Figura 5.3, se puede ver que para $\alpha = 1$, $x_0 = 1$, ambos test dan buenos resultados, obteniéndose una frecuencia de rechazo muy cercana al valor teórico, aún para valores de pendiente m grandes. Para ese mismo valor de x_0 y valores de α mayores, ambos test mantienen los buenos resultados; cabe destacar que para valores de m grandes el test clásico nunca rechaza.

Para $x_0 = 3$, el test clásico sólo es informativo para $m = 0.5$, se puede ver que el nivel del test es afectado por la presencia de outliers de leverage moderado (caso $m = 1, m = 1.5$), ya que la frecuencia de rechazo es alta, y para outliers de leverage alto, pasa a tener

frecuencia de rechazo cero. De esta forma podemos concluir que el test clásico tiene un comportamiento errático aún frente a un solo outlier. Todos estos comportamientos empeoran cuando α aumenta. Por otro lado, el test propuesto muestra muy buenos resultados en todos los casos.

Para $x_0 = 5$, para el test basado en el estimador de máxima verosimilitud se observan los mismos problemas mencionados para el caso $x_0 = 3$. El test propuesto, en cambio, muestra en general buenos resultados, salvo para el caso $\alpha = 1$ y $m = 0.5$ debido a que en este caso el dato atípico reemplazado es intermedio y el test no puede identificarlo como valor anómalo.

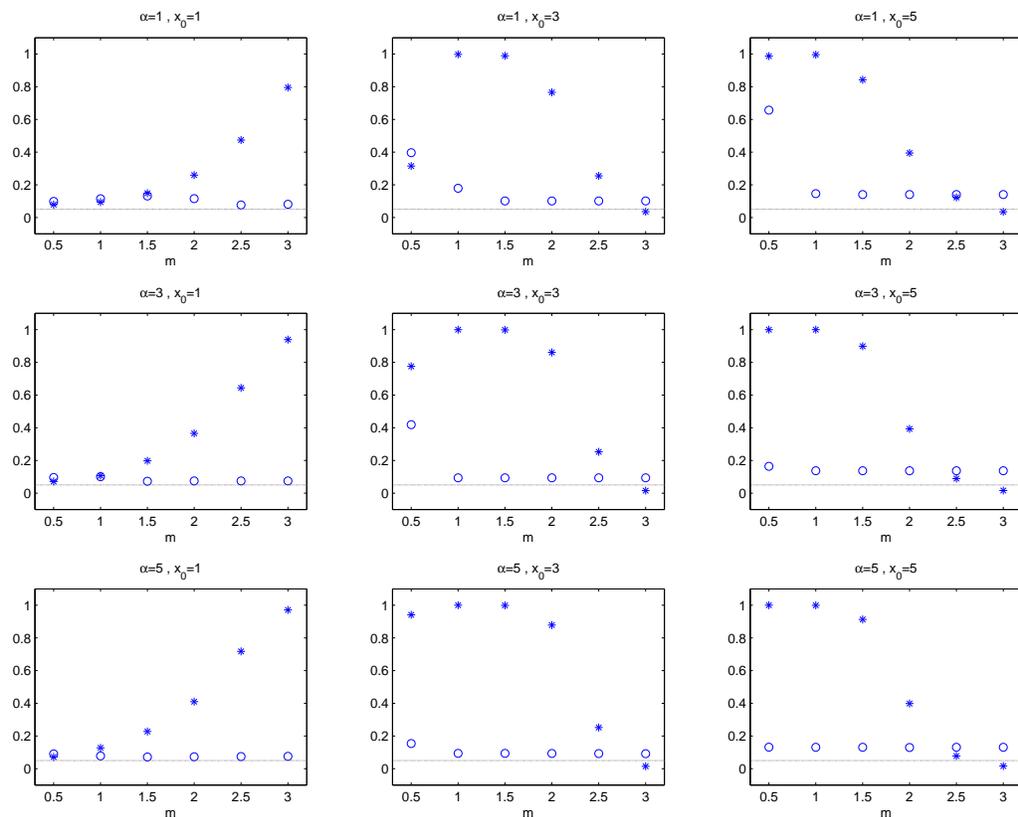


Figura 5.4: Frecuencia observada de rechazo bajo H_0 en el caso C_3 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

La Figura 5.4 muestra que al contaminar con un 3% de outliers con $x_0 = 1$, el test clásico mantiene un nivel aceptable para outliers con pendiente m pequeña, pero la frecuencia de rechazo aumenta

hacia 1 a medida que la pendiente crece para todos los valores de α considerados.

Para $x_0 = 1$, el test propuesto muestra en todos los casos una frecuencia de rechazo bajo H_0 próxima al valor nominal, mostrándose muy estable bajo la contaminación C_3 .

En el caso de $x_0 = 3$, para $\alpha = 1$ y $\alpha = 3$, para $m = 0.5$, el test propuesto muestra cierta inestabilidad, seguramente asociada al hecho de que los outliers introducidos son de difícil detección ya que son de baja atipicidad, como puede verse en los gráficos 5.1 y 5.2.

Cuando $x_0 = 5$ se observan los mismos comportamientos que para $x_0 = 3$, pero para $\alpha \geq 3$, el robusto mantiene un nivel estable.

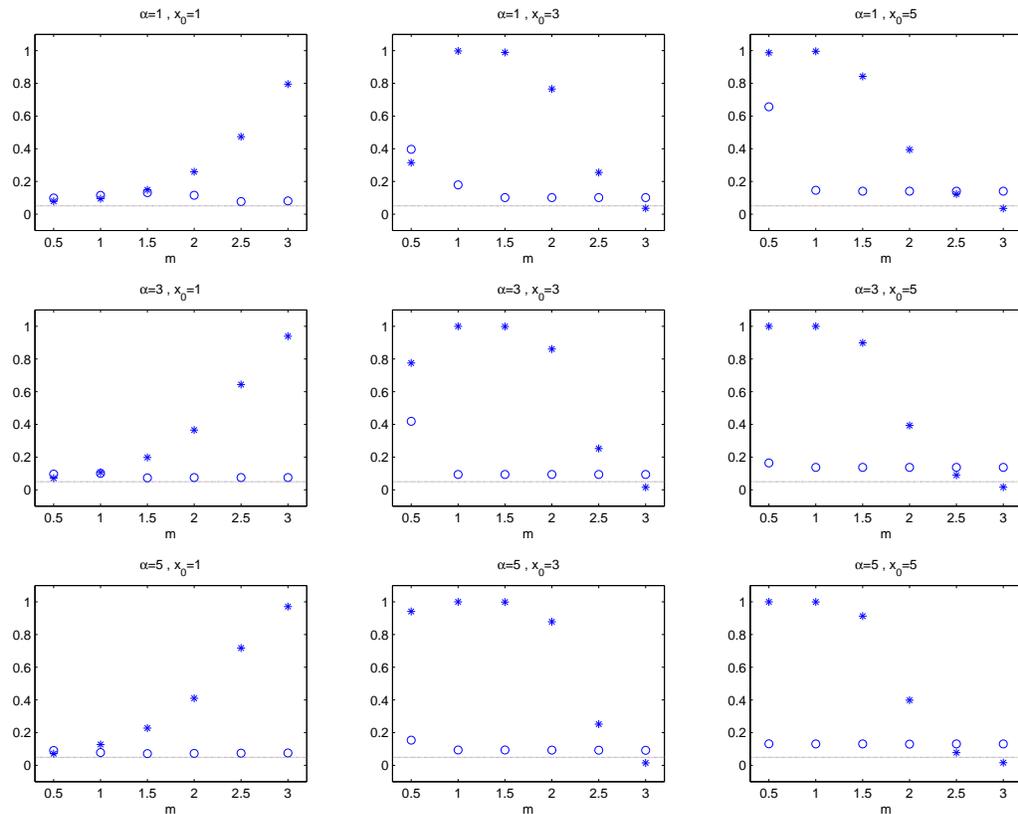


Figura 5.5: Frecuencia observada de rechazo bajo H_0 en el caso C_5 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

Las conclusiones a partir de Figura 5.5 son similares a las anteriores.

En gran parte de los casos estudiados para $\alpha = 3$ y 5 , las pendientes $m = 1$ y $m = 1.5$ representan las situaciones más críticas para el test clásico, mientras que el test propuesto muestra una gran estabilidad de nivel.

5.2. Potencia

El propósito de esta sección es estudiar la potencia alcanzada por el test clásico y el propuesto en muestras de tamaño moderado, tanto contaminadas como sin contaminar.

En este estudio de simulación generamos 5000 muestras independientes de tamaño $n = 100$, con $p = 3$, es decir intercept y dos covariables siguiendo el esquema sin contaminación C_0 y los esquemas de contaminación C_1 , C_3 y C_5 introducidos en la sección anterior correspondientes al caso de pendiente $m = 1$.

Para el estudio de potencia las respuestas y_i se generaron según el modelo LGR dado por

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 + u_i, \quad i = 1, \dots, 100, \quad (5.3)$$

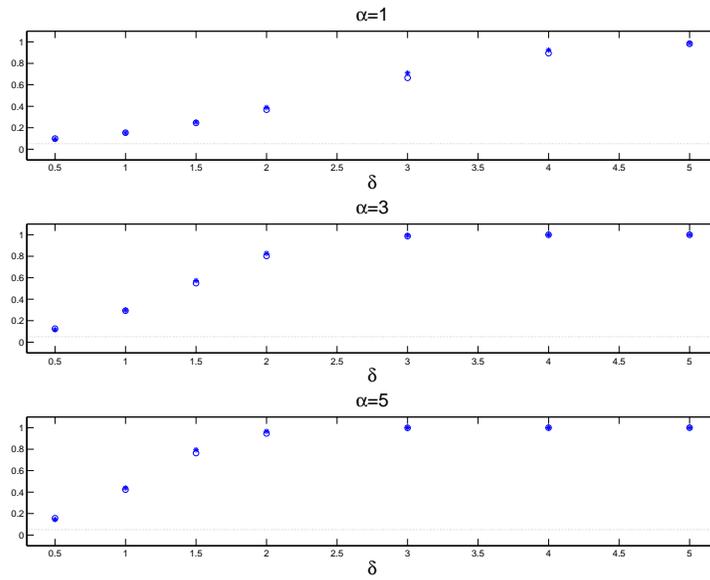
donde u_i tiene distribución $\log \Gamma(\alpha, 1)$ y $\beta' = (0, 0, 0) + (0, \frac{\delta}{\sqrt{100}}, 0)$ para valores de $\delta = 0, 0.5, 1, 1.5, 2, 3, 4$ y 5 .

La hipótesis nula considerada fue $H_0 : \beta = \beta_0$, con $\beta'_0 = (0, 0, 0)$.

- Caso sin contaminar

En el Cuadro 5.2 y la Figura 5.6 se muestran las frecuencias observadas de rechazo en el caso C_0 , cuando el nivel nominal es $\tau = 5\%$. Los dos tests tienen un comportamiento similar para muestras sin contaminar para los tres valores del parámetro de forma α considerados. Notamos que ambos tests son más potentes a medida que crece el valor de α .

		δ	0	0.5	1	1.5	2	3	4	5
$\alpha = 1$	D_{MV}	0.077	0.093	0.152	0.253	0.388	0.710	0.924	0.988	
	D_R	0.075	0.099	0.154	0.245	0.368	0.665	0.895	0.982	
$\alpha = 3$	D_{MV}	0.061	0.116	0.298	0.573	0.827	0.994	1	1	
	D_R	0.077	0.126	0.293	0.549	0.803	0.987	1	1	
$\alpha = 5$	D_{MV}	0.062	0.145	0.439	0.794	0.966	1	1	1	
	D_R	0.080	0.157	0.423	0.764	0.947	0.999	1	1	

Cuadro 5.2: Frecuencia observada de rechazo en el caso C_0 , $\tau = 5\%$ Figura 5.6: Frecuencia observada de rechazo en el caso C_0 , $\tau = 5\%$.

- Caso contaminado

Al igual que en el estudio de nivel de la Sección 5.1, los outliers fueron tomados de la forma $(x_0, 0, mx_0)$ para valores de $x_0 = 1$, $x_0 = 3$ y $x_0 = 5$, y en este caso fijamos la pendiente $m = 1$.

Los resultados son reportados en las Figuras 5.7 a 5.9. En todas ellas la frecuencia observada de rechazo se representa mediante un 'o' para el test robusto y un '*' para el clásico; la línea punteada

representa el nivel nominal 5 %.

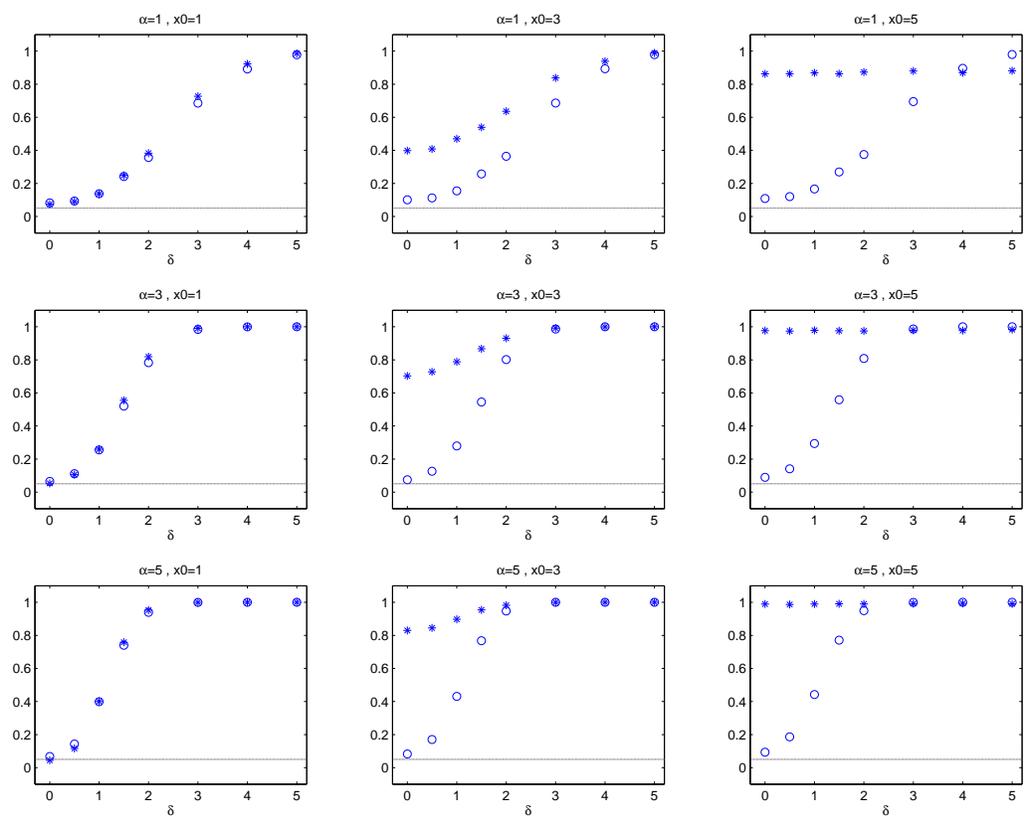


Figura 5.7: Frecuencia observada de rechazo en el caso C_1 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

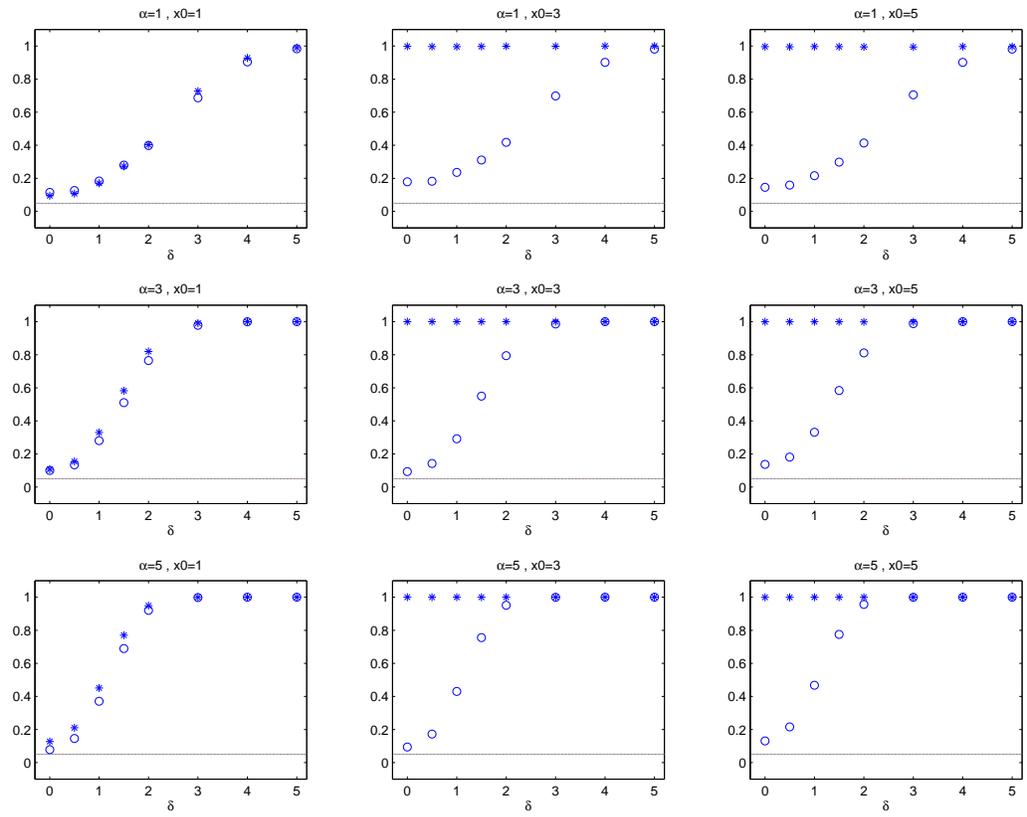


Figura 5.8: Frecuencia observada de rechazo en el caso C_3 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

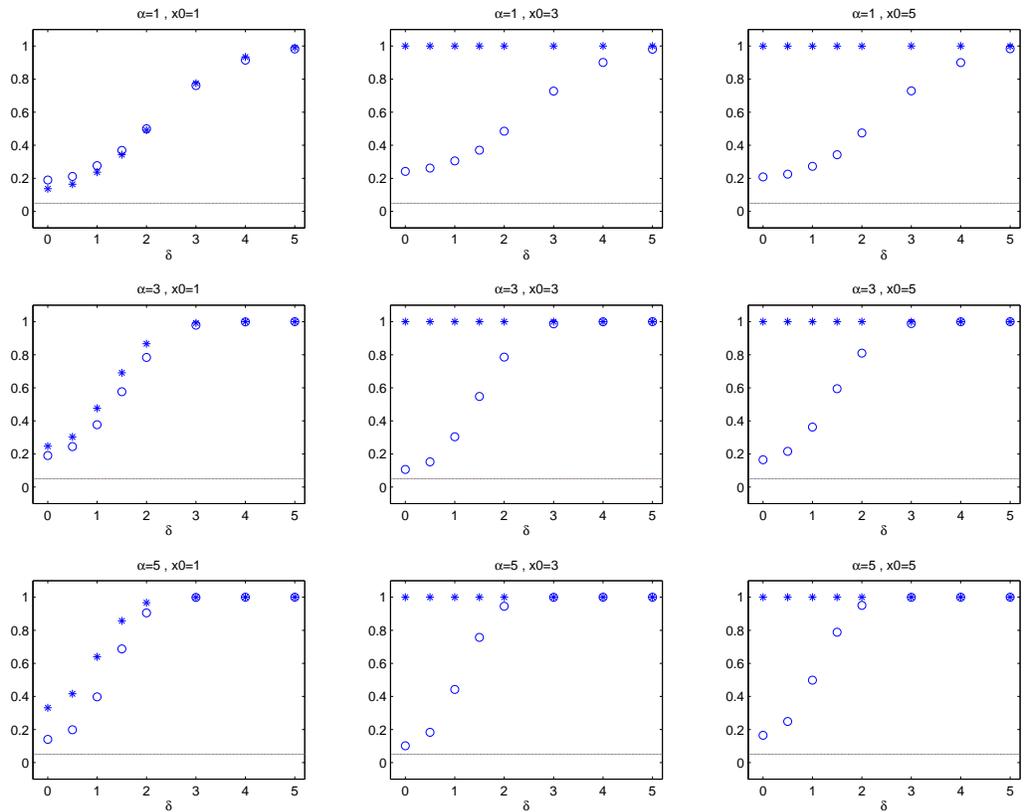


Figura 5.9: Frecuencia observada de rechazo en el caso C_5 , $\tau = 5\%$, 'o' caso robusto y '*' el caso clásico.

Cuando $x_0 = 1$ ambos tests dan resultados muy similares a los obtenidos en las muestras no contaminadas reportados en la Figura 5.6, en especial bajo C_1 y C_3 ; bajo C_5 hay cierta pérdida de potencia por parte del test clásico.

Para $x_0 = 3$ el comportamiento de los tests comienza a diferenciarse, aún ante la presencia de un solo outlier. De hecho, para el caso más benigno que corresponde a $\alpha = 1$ bajo la contaminación C_1 , el test clásico rechaza un 40% de las veces bajo la hipótesis nula, como puede apreciarse en la Figura 5.7. Este comportamiento empeora aún más cuando α crece y cuando aumenta el porcentaje de contaminación, como se observa en los Figuras 5.8 y 5.9, correspondientes a los esquemas de contaminación C_3 y C_5 , respectivamente. En efecto, la potencia del test clásico da constantemente 1 para $x_0 = 3$ tanto bajo C_3 como bajo C_5 , mostrando que el test clásico es no informativo en estas situaciones. Por el contrario, en

todos estos casos el test robusto conserva una potencia similar a la obtenida en muestras sin contaminar.

Para $x_0 = 5$, la función de potencia del test clásico es igual a 1 para cualquier valor de δ , incluso bajo la hipótesis nula, por lo tanto no es informativo, mientras que la potencia del test propuesto es muy estable.

Es decir, para los casos en que $x_0 = 3$ y $x_0 = 5$, el test clásico es no informativo en tanto rechaza siempre, mientras que el test propuesto, si bien puede alejarse levemente del valor teórico bajo la hipótesis nula en algunos casos, muestra buenos resultados mostrando que no se ve afectado por la presencia de datos atípicos.

Capítulo 6

Ejemplos

6.1. Datos de Feigl y Zelen

La leucemia mielógena aguda (LMA) es un cáncer de la médula ósea y la sangre que se caracteriza por el crecimiento incontrolado de los glóbulos blancos inmaduros conocidos como mielocitos. Esta enfermedad es más común en los adultos que en los niños, con un promedio de edad al momento del diagnóstico inicial de más de 65 años.

Feigl y Zelen (1965) dan el tiempo de supervivencia en semanas desde el diagnóstico inicial de 33 pacientes que murieron a causa de esta enfermedad. Para cada paciente, disponen de información adicional sobre dos covariables

- **WBC**: el recuento de glóbulos blancos
- **AG**: una variable factor que clasifica a los pacientes como positivos o negativos dependiendo de la presencia o ausencia de una característica morfológica en los glóbulos blancos.

La variable **WBC** es una herramienta útil de diagnóstico del paciente, en tanto altos valores de la misma parecen estar asociados a condiciones más severas de la enfermedad.

Con el fin de entender como el comportamiento de las variables **WBC** y **AG** impacta sobre el tiempo de supervivencia, y asimismo poder entender y describir la posible relación funcional entre las mismas, formulamos un modelo de predicción.

Los datos fueron ajustados mediante el modelo log-Gamma dado por

$$\log(y_i) = \beta_1 WBC_i + \beta_2 AG_i + \beta_3 + u_i,$$

donde u_i tiene distribución $\log \Gamma(\alpha, 1)$.

En la primera y segunda columnas del Cuadro 6.1 se encuentran los coeficientes estimados y sus errores estándar (SE) para el *MLE* y el *MM*-estimador respectivamente para los datos completos. En la Figura 6.1 se muestra el QQ-plot de los residuos correspondientes a la estimación con el *MM*-estimador, éste muestra que los residuos siguen bastante bien el supuesto de una distribución log-Gamma y revela cuatro claros valores extremos. Estos corresponden a pacientes con un recuento de células blancas muy alto que sobrevivieron más de lo esperado. El Cuadro 6.1 asimismo contiene en la tercera columna los coeficientes estimados y errores estándar correspondientes al *MLE*, después de quitar los cuatro valores extremos detectados mediante la Figura 6.1.

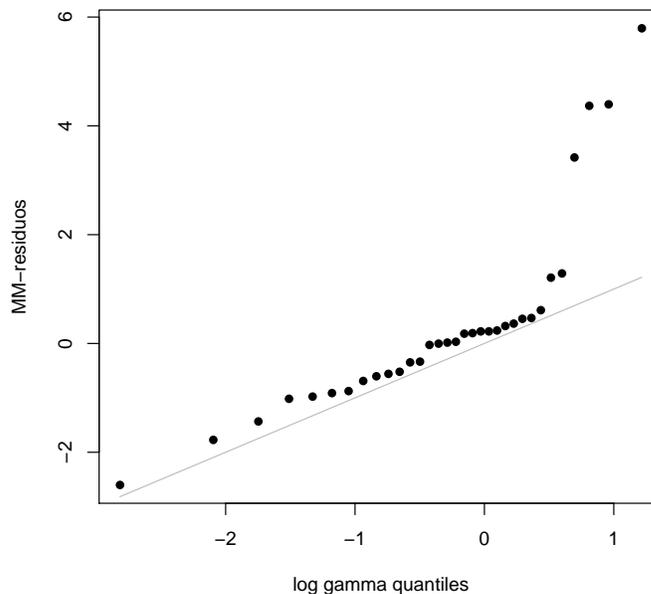


Figura 6.1: QQ-plot de los *MM*-residuos.

En el Cuadro 6.1, podemos observar la diferencia en los co-

Variable	<i>MLE</i>		<i>MM</i>		<i>MLE</i>	
	Datos completos		Datos Completos		4 casos removidos	
	$\hat{\beta}_i$	SE	$\hat{\beta}_i$	SE	$\hat{\beta}_i$	SE
WBC(/100)	-0.0006	0.0004	-0.0051	0.0003	-0.0051	0.0004
AG(0-1)	-1.10	0.34	-1.81	0.27	-1.57	0.27
Intercept	4.23	0.27	4.85	0.22	4.8	0.21
α	0.97		1.60		1.81	

Cuadro 6.1: Coeficientes estimados y error estándar.

eficientes estimados al utilizar el *MM*-estimador propuesto y el *MLE*. La diferencia más importante se observa en el valor de $\hat{\beta}_1$, el coeficiente que acompaña a la variable **WBC**. El Cuadro 6.1 muestra que después de la supresión de los cuatro valores atípicos, los coeficientes estimados al utilizar el *MLE* son bastante similares a los obtenidos con el *MM*-estimador.

En segunda instancia, con el fin de estudiar la significación de cada una de las covariables presentes se realizaron test individuales con un nivel individual de 5%. En cada caso se testeó $H_0^* : \beta_{(i)} = 0$, $1 \leq i \leq 3$ y se comparó con el percentil $\chi_{1, 0.05}^2 = 3.841$. El Cuadro 6.2 muestra el valor del estadístico obtenido en cada caso, mientras que en el Cuadro 6.3 se hallan las decisiones a las que conducen los tests de tipo Wald clásicos y robustos.

H_0^*	<i>MLE</i>	<i>MM</i>	<i>MLE</i>
	Datos completos	Datos Completos	4 casos removidos
	D_i	D_i	D_i
$WBC(/1000) = 0$	2.67	254.7	140.7
$AG = 0$	18.79	78.8	58.5
$Intercept = 0$	571.30	1162.3	1211.4

Cuadro 6.2: Valor del estadístico de Wald al testear $H_0^* : \beta_{(i)} = 0$, $1 \leq i \leq 3$.

El Cuadro 6.3 muestra que, contrariamente a lo esperado, para el test clásico la variable **WBC** no es significativa, ya que el test no rechaza la hipótesis nula $H_0^* : \beta_{(1)} = 0$. Por otro lado, el test propuesto lleva a la conclusión opuesta. Después de la supresión de los cuatro valores atípicos detectados, el test clásico nos llevan

	<i>MLE</i>	<i>MM</i>	<i>MLE</i>
	Datos completos	Datos Completos	4 casos removidos
H_0^*	Respuesta	Respuesta	Respuesta
$WBC(/1000) = 0$	no rechaza	rechaza	rechaza
$AG = 0$	rechaza	rechaza	rechaza
$Intercept = 0$	rechaza	rechaza	rechaza

Cuadro 6.3: Decisión al testear $H_0^* : \beta_{(i)} = 0, 1 \leq i \leq 3$ al 5%.

a concluir que esta variable es significativa, arrojando el mismo resultado que el obtenido a partir del test robusto. Esto coincide con lo esperado por los médicos, quienes intuían que la cantidad de glóbulos blancos es una variable muy importante a considerar en la predicción del tiempo de supervivencia de los pacientes.

6.2. Enfermedades Vasculares

Las enfermedades vasculares constituyen la principal causa de muerte en Portugal. Usualmente los factores que aumentan el riesgo de sufrir infartos y accidentes cerebro vasculares (ACV) son el hipercolesterol, la hipertensión arterial y el tabaquismo.

Hasta hace un tiempo, un nivel de homocisteína elevado en sangre se consideraba un dato meramente anecdótico, sin embargo estudios científicos recientes han mostrado que incluso los niveles moderadamente elevados de homocisteína aumentan el riesgo de enfermedad de las arterias coronarias, cerebrales, periféricas y de la muerte cardiovascular.

La hiperhomocisteinemia, un grupo de enfermedades metabólicas que se caracterizan por presentar un nivel elevado del aminoácido homocisteína en plasma, se traduce en valores elevados de la homocisteinemia basal (HBP) y en la homocisteinemia después de la sobrecarga de la metionina oral (HSP).

Amaral-Turkman y Silva (2000) presentan un conjunto de datos analizados en Teles (1995) con los que se realizó una investigación con el propósito de averiguar qué variables influyen significativamente en los valores de la variable HBP. Fueron observados 127

individuos que sufrieron un ataque al corazón o un derrame cerebral entre los 26 y los 56 años de edad.

En este análisis, consideramos para cada individuo, la variable de respuesta **HBP** y las covariables:

- **Sexo**: 0 (femenino), 1 (masculino).
- **Histfa**: 0 (sin antecedentes familiares), 1 (con antecedentes familiares).
- **Creat**: creatinina.
- **Falc**: fosfato alcalino.
- **GGT**: gamma-glutamyl-transpeptidasa.
- **Pad**: presión arterial diastólica.
- **Fumar**: 0 (no fumador), 1 (fumador), 2 (ex fumador).

Para representar la covariable **Fumar** se utilizaron dos covariables **F1** y **F2**.

Los datos fueron ajustados mediante el modelo log-Gamma dado por

$$\log(y_i) = \beta_1 \text{Sexo}_i + \beta_2 \text{Histfa}_i + \beta_3 \text{Creat}_i + \beta_4 \text{Falc}_i + \beta_5 \text{GGT}_i + \beta_6 \text{Pad}_i + \beta_7 \text{F1}_i + \beta_8 \text{F2}_i + \beta_9 + u_i,$$

donde u_i tiene distribución $\log \Gamma(\alpha, 1)$.

Al igual que en el ejemplo anterior, en la primera y segunda columna del Cuadro 6.4 se encuentran los coeficientes estimados y sus errores estándar para el *MLE* y el *MM*-estimador respectivamente para los datos completos.

En la Figura 6.2 se muestra el QQ-plot de los residuos correspondientes a la estimación con el *MM*-estimador, este muestra que los residuos siguen una distribución log-Gamma y revela 4 valores extremos, 3 outliers importantes y uno moderado. Estos corresponden a pacientes que presentan valores elevados de la homocisteinemia basal (HBP).

El Cuadro 6.4 contiene en la tercera columna los coeficientes estimados y los errores estándar correspondientes al *MLE*, después de quitar los cuatro valores extremos detectados mediante la Figura 6.2.

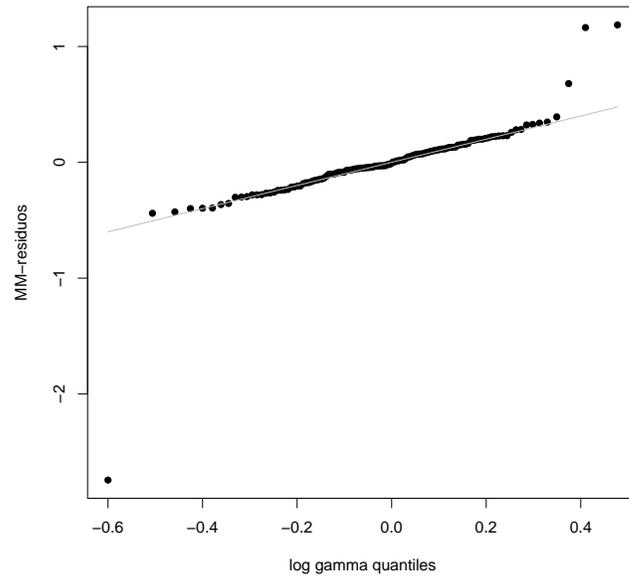


Figura 6.2: QQ-plot de los *MM*-residuos.

Variable	<i>MLE</i>		<i>MM</i>		<i>MLE</i>	
	Datos completos		Datos Completos		4 casos removidos	
	$\hat{\beta}_i$	SE	$\hat{\beta}_i$	SE	$\hat{\beta}_i$	SE
Sexo	0.2918	0.0716	0.2195	0.0516	0.2026	0.0514
Histfa	-0.0997	0.0502	-0.0279	0.0362	-0.0161	0.0352
Creat	0.1009	0.0249	0.3514	0.0179	0.3695	0.0940
Falc	0.0032	0.0007	0.0017	0.0005	0.0017	0.0005
GGT	-0.0016	0.0006	-0.0011	0.0005	-0.0011	0.0004
Pad	-0.0055	0.0007	-0.0027	0.0005	-0.0029	0.0005
Fumar1	0.2387	0.0711	0.0555	0.0512	0.0685	0.0514
Fumar2	-0.0064	0.0555	-0.0168	0.0400	-0.0005	0.0383
Intercept	2.2602	0.2502	1.8894	0.1803	1.8734	0.1970
α	12.97		24.36		27.35	

Cuadro 6.4: Coeficientes estimados y SE.

En segunda instancia, con el fin de estudiar la significación de

cada una de las covariables presentes se realizaron test individuales con un nivel individual de 5%. En cada caso se testeó $H_0^* : \beta_{(i)} = 0$, $1 \leq i \leq 9$ y se comparó con percentil $\chi_{1, 0.05}^2 = 3.841$. El Cuadro 6.5 muestra el valor del estadístico obtenido en cada caso, mientras que en el Cuadro 6.6 se hallan las decisiones a las que conducen los tests de tipo Wald clásicos y robustos.

Vale notar que los valores obtenidos mediante el *MM*-estimador son bastante parecidos a los obtenidos con el método de máxima verosimilitud una vez que las 4 observaciones atípicas han sido removidas. Es más interesante aún comprobar en el Cuadro 6.6 que en los nueve tests individuales siempre que el test clásico basado en la muestra completa difiere en su conclusión del correspondiente test basado en el estadístico robusto, éste último coincide con el test clásico aplicado a la muestra sin los cuatro valores atípicos, mostrando una vez más la estabilidad del test robusto frente a valores anormales en la muestra.

	<i>MLE</i>	<i>MM</i>	<i>MLE</i>
	Datos completos	Datos Completos	4 casos removidos
H_0^*	D_i	D_i	D_i
Sexo	118.16	126.5	115.70
Histfa	9.41	1.39	0.51
Creat	36.77	843.92	481.49
Falc	124.42	64.34	75.28
GGT	14.65	13.51	4.45
Pad	385.22	180.58	217.05
Fumar1	17	1.73	0.0004
Fumar2	0.03	0.40	2.56
Intercept	8414.8	11129	11808

Cuadro 6.5: Valor del estadístico de Wald al testear $H_0^* : \beta_{(i)} = 0$, $1 \leq i \leq 9$.

	<i>MLE</i>	<i>MM</i>	<i>MLE</i>
	Datos completos	Datos Completos	4 casos removidos
H_0^*	Respuesta	Respuesta	Respuesta
Sexo	rechaza	rechaza	rechaza
Histfa	rechaza	no rechaza	no rechaza
Creat	rechaza	rechaza	rechaza
Falc	rechaza	rechaza	rechaza
GGT	rechaza	rechaza	rechaza
Pad	rechaza	rechaza	rechaza
Fumar1	rechaza	no rechaza	no rechaza
Fumar2	no rechaza	no rechaza	no rechaza
Intercept	rechaza	rechaza	rechaza

Cuadro 6.6: Decisión al testear $H_0^* : \beta_{(i)} = 0, 1 \leq i \leq 9$ al 5%.

Bibliografía

- [1] Amaral-Turkman, M.A. y Silva, G.L. (2000). *Modelos Lineares Generalizados - da teoria à prática*. Edições SPE, Lisboa, Portugal.
- [2] Bianco, A., Garcia Ben, M. y Yohai, V.(2005). Robust estimation for linear regression with asymmetric errors. *The Canadian Journal of Statistics*, **33**, No.4, 1-18.
- [3] Cantoni, E. y Ronchetti, E.(2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022-1930.
- [4] Donoho, D.y Huber, P. (1983). The notion of breakdown point. *In A Festschrift for Erich L. Lehmann (Bickel P., Docksum K. y Hodges J., Jr., eds.)*, Wadsworth, Belmont, California, 157-184.
- [5] Feigl, P. y Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826-838.
- [6] Heritier, S. y Ronchetti, E.(1994). Robust Bounded-Influence Tests in General Parametric Models. *Journal of the American Statistical Association*, **89**, No. 427, 897-904.
- [7] Künsch, H., Stefanski, L. y Carroll, R.(1989). Conditionally unbiased bounded-influence estimation in general regression models with applications to generalized linear models. *Journal of the American Statistical Association*, **84**, 460-466.
- [8] Marazzi, A. y Ruffieux, C. (1996). Implementing M-estimators of the gamma distribution. In Robust Statistics. *Data Analysis*

- and Computer Intensive Methods, in Honor of Peter Huber's 60th Birthday* (H. Rieder, ed.), *Lecture Notes in Statistics 109*, Springer, New York, 277-297.
- [9] Maronna, R., Martin, R. y Yohai, V. (2006). *Robust Statistics—Theory and Methods*. John Wiley and Sons Ltd.
- [10] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. (2nd ed.) London: Chapman and Hall.
- [11] Rousseeuw, P. y Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- [12] Stefanski L., Carroll, R. y Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413-424.
- [13] Teles, J.M.V. (1995). *Modelos Lineares Generalizados - Uma Aplicação à Medicina*. Tesis de Maestría en Estadística e Investigación Operativa. Facultad de Ciencias de la Universidad de Lisboa, Portugal.
- [14] Yohai, V. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, **15**, No. 2, 642-656
- [15] Yohai V, y Maronna, R. (1979). Asymptotic behavior of M -estimates for the linear model. *The Annals of Statistics*, **7**, 258-268