



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MEDELLÍN
FACULTAD CIENCIAS
ESCUELA DE ESTADÍSTICA

Una Estimación Robusta y no paramétrica de la Transformación de Box y Cox para Series de Tiempo

Dasy Andrea Calle Correa

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2015

Una Estimación Robusta y no paramétrica de la Transformación de Box y Cox para Series de Tiempo

Dasy Andrea Calle Correa

Tesis presentado como requisito parcial para optar al título de:
Magister en Ciencias - Estadística

Director:
Elkin Castaño Vélez, M.Sc.

Línea de Investigación:
Series de Tiempo
Grupo de Investigación en Estadística Universidad Nacional de Colombia, Sede Medellín

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2015

Dedicatoria

A mis padres María Teresa y Jorge Alberto

Agradecimientos

Agradezco a todas las personas que me ayudaron a cumplir este sueño. A todos los profesores que compartieron su conocimiento, especialmente al profesor Elkin Castaño que estuvo presente en todo el proceso de mi formación como profesional.

Resumen

En el análisis de series de tiempo estacionarias, es frecuente encontrarse que la varianza de la serie no es constante, siendo necesario en estos casos transformar la serie utilizando la familia de transformaciones introducida por Box y Cox (1964), donde se busca una transformación de potencia que permita estabilizar la varianza de la serie. Sin embargo, varios autores han estudiado y demostrado que la familia de transformaciones de Box y Cox parece no ser muy adecuada ni robusta cuando hay existencia de observaciones atípicas en la serie, la presencia de estas observaciones distorsiona la estimación del parámetro de transformación de Box y Cox (λ).

En este trabajo se plantea estudiar una propuesta para la estimación robusta del parámetro λ de la familia de transformaciones de box y cox que sea robusto ante al presencia de observaciones atípicas y tenga en cuenta el efecto de estas.

Palabras claves: Series de tiempo, transformación, datos atípicos, varianza.

Abstract

In the stationary time series analysis it is frequent to encounter that the variance of the series is not constant. In these cases it is necessary to transform the series by using the family of transformations introduced by Box and Cox (1964), where it is sought a power transformation that allows stabilizing the variance of the series. However, several authors have studied and proved that the family of transformations Box and Cox seem not to be very adequate nor robust when there is presence of atypical observations in the series. The presence of these observations distorts the estimation of the parameter of transformation Box and Cox (λ).

In this work it is set out the study of a proposal to the robust estimation of the parameter λ of the family of transformations Box and Cox, which is robust facing the presence of atypical observations and takes in account the effect of these..

Keywords: Time series, transformation, outliers, variance.

Contenido

Lista de Tablas	xiii
Lista de Figuras	xv
1. Introducción	1
1.1. Introducción	1
2. Marco Teórico	5
2.1. Series de tiempo	5
2.1.1. Series de tiempo estacionarias	5
2.1.2. Series de tiempo no estacionarias	8
2.2. Familia de transformaciones de Box y Cox	9
2.2.1. Transformación de Box y Cox en la Regresión Lineal	9
2.2.2. Transformación de Box y Cox en Series de Tiempo	10
2.3. Observaciones atípicas	10
2.3.1. Observaciones atípicas (outliers) en series de tiempo	11
3. Metodología propuesta	13
3.1. Transformación de Box y Cox	13
3.2. Transformación de Box y Cox ante la presencia de observaciones atípicas . .	15
3.3. Procedimiento propuesto	17
3.4. Error estándar de la transformación estimada	19
4. Experimento vía simulación	21
4.1. Procedimiento propuesto condicional	21
4.2. Procedimiento propuesto incondicional	22
5. Análisis de resultados	39
6. Conclusiones y recomendaciones	41
6.1. Conclusiones	41
6.2. Recomendaciones	42
A. Algoritmos y simulaciones	43

Bibliografía

47

Lista de Tablas

3-1. Resultados datos simulados con $\lambda = 0,25$	16
3-2. Resultados datos simulados con $\lambda = 0,5$	17
3-3. Resultados datos simulados con $\lambda = 1$	17
4-1. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,25 magnitud 8σ . .	23
4-2. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,25 magnitud 10σ . .	24
4-3. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,50 magnitud 8σ . .	24
4-4. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,50 magnitud 10σ .	25
4-5. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,75 magnitud 8σ . .	25
4-6. Resultados para $\lambda = 0,25$ contaminada en el percentil 0,75 magnitud 10σ . .	26
4-7. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,25 magnitud 8σ . . .	28
4-8. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,25 magnitud 10σ . .	28
4-9. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,5 magnitud 8σ . . .	29
4-10. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,5 magnitud 10σ . . .	29
4-11. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,75 magnitud 8σ . . .	30
4-12. Resultados para $\lambda = 0,5$ contaminada en el percentil 0,75 magnitud 10σ . .	30
4-13. Resultados para $\lambda = 1$ contaminada en el percentil 0,25 magnitud 8σ	32
4-14. Resultados para $\lambda = 1$ contaminada en el percentil 0,25 magnitud 10σ . . .	33
4-15. Resultados para $\lambda = 1$ contaminada en el percentil 0,5 magnitud 8σ	33
4-16. Resultados para $\lambda = 1$ contaminada en el percentil 0,5 magnitud 10σ	34
4-17. Resultados para $\lambda = 1$ contaminada en el percentil 0,75 magnitud 8σ	34
4-18. Resultados para $\lambda = 1$ contaminada en el percentil 0,75 magnitud 10σ . . .	35

Lista de Figuras

3-1. Modelo sin presencia de observaciones atípicas	14
3-2. Parámetro λ cuando no hay presencia de observaciones atípicas	14
3-3. Modelo con presencia de observaciones atípicas	15
3-4. Parámetro λ cuando hay presencia de observaciones atípicas	16
4-1. Resultados para $\lambda = 0,25$ contaminada en el percentil 0.25	26
4-2. Resultados para $\lambda = 0,25$ contaminada en el percentil 0.5	27
4-3. Resultados para $\lambda = 0,25$ contaminada en el percentil 0.75	27
4-4. Resultados para $\lambda = 0,5$ contaminada en el percentil 0.25	31
4-5. Resultados para $\lambda = 0,5$ contaminada en el percentil 0.5	31
4-6. Resultados para $\lambda = 0,5$ contaminada en el percentil 0.75	32
4-7. Resultados para $\lambda = 1$ contaminada en el percentil 0.25	35
4-8. Resultados para $\lambda = 1$ contaminada en el percentil 0.5	36
4-9. Resultados para $\lambda = 1$ contaminada en el percentil 0.75	36

1. Introducción

1.1. Introducción

En el análisis de series de tiempo estacionarias, es común encontrarse que la varianza del proceso no es constante, llevándonos a buscar metodologías para poder estabilizar la varianza de la serie, para esto, frecuentemente se emplea una transformación de Box y Cox, Wei (1994)[26], siendo posible estabilizar la varianza de la serie Z_t por medio de la familia de transformaciones introducida por Box y Cox (1964) [5].

La familia de transformaciones Box y Cox (1964)[5] es un método tradicional y conocido para transformar la variable dependiente de la regresión lineal, no obstante, el parámetro de transformación es válido si los supuestos son satisfechos, características que no necesariamente son cumplidas en los casos donde existe presencia de observaciones atípicas.

Durante los últimos 30 años, varios autores como Atkinson y Shephard (1996)[2], Atkinson y Riani (2000)[1], Riani (2009)[23], Granger y Newbold (1976)[14], Lenk y Tsai (1990)[16], Tsai y Wu (1992)[25], Pankratz (1991)[20] y Guerrero (1993)[15] han estudiado esta problemática, demostrando que la estimación del parámetro de transformación λ se ve gravemente afectado cuando hay observaciones atípicas presentes en la muestra, proponiendo entonces metodologías robustas de identificación, diagnósticos y eliminación de efectos de observaciones atípicas.

Algunos autores como Marazzi y Yohai (2004)[18] presentaron una nueva clase de estimación para el caso de regresión simple, donde los estimadores son robustos y consistentes incluso bajo no normalidad y no homocedasticidad, la estimación se basa en la minimización de una medida de autocorrelación entre los residuos con respecto a una estimación robusta del coeficiente de regresión. Esta medida es una reminiscencia de una propuesta de Maravall (1983)[17] para la detección de no linealidad en series de tiempo.

Carroll y Ruppert (1988)[7] propusieron otro enfoque limitando la influencia de las observaciones atípicas en las estimaciones, basados en un modelo homocedástico. Bickel y Doksum (1981)[4] también muestran un método robusto para estimar el parámetro de transformación basado en desviaciones absolutas.

Riani (2009)[23] presenta una prueba de puntuación basada en un algoritmo de búsqueda hacia adelante en la forma sugerida por Atkinson y Riani (2000)[1] y se extendió a series de tiempo por Riani (2004)[22], este algoritmo es a la vez eficiente y robusto; es eficiente ya que hace uso de la verosimilitud gaussiana y es robusto debido a que los valores atípicos entran en el último paso del procedimiento, y su efecto en las estadísticas de interés es claramente representado.

Castaño (1995, 2011)[8] presenta un procedimiento alternativo no paramétrico y robusto que permite obtener una estimación del parámetro λ en la familia de transformaciones de potencia de Box y Cox cuando existen observaciones atípicas en la variable dependiente para un modelo de regresión por medio de búsqueda directa usando regresión robusta de mínima desviación absoluta LAD (Least Absolute Deviation).

Se propone extender la propuesta de Castaño (1995, 2011)[8] para series de tiempo debido a que frecuentemente en el análisis de series de tiempo estacionarias es necesario encontrar una transformación adecuada para estabilizar la varianza de la serie que tenga en cuenta la influencia de observaciones atípicas.

En el análisis de series de tiempo estacionarias, la serie está caracterizada por su media, su varianza, sus autocorrelaciones y sus autocorrelaciones parciales. Sin embargo, en la práctica muchas series que deben ser analizadas no son estacionarias, en consecuencia los resultados para series de tiempo se vuelven inválidos y no ciertos, por esta razón se debe buscar metodologías alternativas o adaptar métodos existentes de tal forma que las condiciones se cumplan. La no estacionaridad de un proceso puede ocurrir en diferentes formas: no estacionaridad en media, en varianza y en covarianza.

En la práctica es muy común que los procesos estudiados no sean estacionarios en varianza, en estos casos es posible estacionarizarla usando la familia de transformaciones introducida por Box y Cox (1964)[5], la cual puede estabilizar su varianza con el fin de mejorar la homogeneidad de los errores. Infortunadamente la búsqueda de la transformación adecuada puede verse afectada cuando en la serie existe presencia de observaciones atípicas. Varios autores han demostrando que la estimación del parámetro de transformación λ se ve gravemente afectado cuando hay observaciones atípicas presentes en la muestra, llevándonos a buscar nuevas estrategias y metodologías para tratar estos casos.

En consecuencia, es importante realizar una debida identificación del modelo, por este motivo, se debe tener precaución en la decisión de transformar o no la serie, teniendo en cuenta que ante la presencia de observaciones atípicas la estimación del parámetro de transformación será afectada llevándonos a tomar decisiones equivocadas, perdiendo confiabilidad en los resultados. La etapa de identificación en el análisis de series de tiempo es fundamental

para poder realizar un buen diagnóstico y así el objetivo del análisis sea cumplido a cabalidad.

En el Capítulo 2 están contenidos los conceptos básicos que serán empleados en este trabajo. En el Capítulo 3 se expone la estimación robusta y no paramétrica de la transformación de Box y Cox para series de tiempo propuesta. En el Capítulo 4 se presenta la propuesta de la estimación vía simulación y se presentan las comparaciones de los resultados obtenidos. En el Capítulo 5 se encuentra el análisis de los resultados recopilados en este trabajo. En el Capítulo 6 se exponen las conclusiones y se proponen algunas recomendaciones. Al final del documento se encuentran los programas utilizados para el desarrollo de esta investigación.

2. Marco Teórico

2.1. Series de tiempo

Una serie de tiempo o serie temporal es una colección de observaciones tomadas a lo largo del tiempo, con la cual se puede describir, explicar, predecir y controlar algún proceso. Las observaciones están ordenadas respecto al tiempo, equiespaciadas y sucesivas Wei (1994)[26].

Las series de tiempo pueden encontrarse en una gran variedad de campos como: agricultura, economía, ingeniería, geofísica, medicina, meteorología, control de calidad, ciencias sociales. La lista de áreas en las que se observa y estudia las series de tiempo es interminable.

El estudio de series de tiempo tiene varios objetivos, como la comprensión de la generación de los datos de la serie o pronosticar valores futuros y control óptimo de sistemas. La naturaleza intrínseca de las series de tiempo es que las observaciones son dependientes o correlacionadas, y por lo tanto el orden es importante. Consecuentemente los procedimientos y técnicas que dependen de asumir independencia no son aplicables y son necesarios otros métodos diferentes.

2.1.1. Series de tiempo estacionarias

Formalmente, una serie de tiempo es una realización del proceso estocástico y corresponde a una observación del proceso. Un proceso estocástico es llamado estacionario de orden n en distribución si la función de distribución conjunta n -dimensional F es invariante en el tiempo. Es decir, si:

$$F(z_{t_1}, z_{t_2}, \dots, z_{t_n}) = F(z_{t_1+k}, z_{t_2+k}, \dots, z_{t_n+k}) \quad (2-1)$$

para cualquier n -tupla (t_1, t_2, \dots, t_n) y k enteros.

Un proceso estocástico es llamado estrictamente estacionario, fuertemente estacionario o completamente estacionario si (2-1) es cierto para todo $n = 1, 2, \dots$

En la práctica, generalmente es muy difícil probar si un proceso estocástico es estrictamente estacionario, y en su lugar se trata de caracterizar los procesos estocásticos en términos de

sus momentos, cuyas propiedades se pueden verificar más fácilmente.

Un proceso es llamado débilmente estacionario de orden n , si todos sus momentos conjuntos hasta orden n son finitos e invariantes en el tiempo.

Un proceso débilmente estacionario de segundo orden tendrá media y varianza constante y sus funciones de covarianza y correlación solamente dependerán del número de periodos que separan los términos del proceso. Esta clase de proceso también es llamado proceso estacionario en sentido amplio o proceso estacionario en covarianza o, simplemente estacionario.

En la práctica generalmente se trabaja con procesos estacionarios en covarianza. Este es un supuesto mucho menos restrictivo que la estacionaridad estricta y más fácil de probar en la práctica.

Para un proceso estrictamente estacionario cuyos momentos de segundo orden existen, puesto que la función de distribución es la misma para todo t , se cumple que:

- La función de medias del proceso: $\mu_t = \mu$, constante
- La función de varianzas del proceso: $\sigma_t^2 = \sigma^2$, constante
- La función de covarianzas entre Z_{t_1} y Z_{t_2} es

$$\text{cov}(Z_{t_1-k}, Z_{t_2+k}) = \text{cov}(Z_{t_1}, Z_{t_2+k}) = \gamma_k \quad (2-2)$$

es decir, solamente depende del número de periodos que separan a Z_{t_1} y Z_{t_2}

- La función de correlación entre Z_{t_1} y Z_{t_2} es

$$\text{corr}(Z_{t_1-k}, Z_{t_2+k}) = \text{corr}(Z_{t_1}, Z_{t_2+k}) = \rho_k \quad (2-3)$$

puesto que solamente depende del número de periodos que separan Z_{t_1} y Z_{t_2}

Los modelos para series de tiempo univariadas estacionarias son:

- El proceso autorregresivo de orden p , AR(p)

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2-4)$$

$$\phi(B)Z_t = \theta_0 + a_t \quad (2-5)$$

donde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ y $\{a_t\}$ es un proceso de ruido blanco de media cero, el proceso es siempre invertible ya que $\sum_{j=1}^p |\phi_j| < \infty$, para que el proceso sea estacionario es necesario que las raíces de $\phi(B) = 0$ caigan fuera del círculo unitario.

- El proceso medias móviles de orden q, MA(q)

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2-6)$$

$$Z_t = \mu + \theta(B)a_t \quad (2-7)$$

donde $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ y $\{a_t\}$ es un proceso de ruido blanco de media cero, el proceso es siempre invertible ya que $\sum_{j=1}^q |\theta_j| < \infty$, para que el proceso sea estacionario es necesario que las raíces de $\theta(B) = 0$ caigan fuera del círculo unitario.

- El proceso mixto autorregresivo y de medias móviles, ARMA(p,q)

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2-8)$$

$$\phi(B)Z_t = \theta_0 + \theta(B)a_t \quad (2-9)$$

donde

$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ es el operador autorregresivo estacionario.

$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ es el operador medias móviles invertible.

No existen factores comunes entre $\phi(B)$ y $\theta(B)$, y $\{a_t\}$ es un proceso de ruido blanco de media cero.

Estacionaridad: las raíces de $\phi(B) = 0$ deben caer fuera del círculo unitario.

Invertibilidad: las raíces de $\theta(B) = 0$ deben caer fuera del círculo unitario.

2.1.2. Series de tiempo no estacionarias

Las series de tiempo descritas anteriormente son procesos estacionarios, sin embargo, muchas series de tiempo, en particular las derivadas de las áreas económicas y de negocios, son no estacionarias. La no estacionaridad en series de tiempo puede ocurrir de muchas maneras diferentes: no estacionaria en media, no estacionaria en varianza, no estacionaria en covarianza, no estacionaria ni en varianza ni en covarianza.

Es por esto que los procedimientos conocidos para trabajar series de tiempo estacionarias no son ciertos para estos casos, sin embargo, algunos procedimientos se pueden adaptar de tal forma que los resultados sean válidos.

Series de tiempo no estacionarias en media

En una serie no estacionaria en media, en general μ_t depende del tiempo, siendo μ_t una función con tendencias determinísticas o con tendencias estocásticas. El tipo de comportamiento estocástico ha sido denominado no estacionaridad homogénea y en estas series su comportamiento local es independiente de su nivel. Las series que exhiben este comportamiento pueden ser convertidas a estacionarias por medio de la diferenciación y d es el menor número de veces que hay que diferenciar la serie para volverla estacionaria. A esta clase de procesos se les llama procesos estacionarios en diferencias o procesos DS (Difference Stationary).

Los modelos para series de tiempo no estacionarias homogéneas se conocen como modelos autorregresivos integrados de media móvil llamados $ARIMA(p, d, q)$, definido así:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (2-10)$$

donde

$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ es el operador autorregresivo estacionario.

$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ es el operador medias móviles invertible.

No existen factores comunes entre $\phi_p(B)$, $\theta_q(B)$, a_t es ruido blanco de media cero, $(1 - B)^d$ corresponde a la d -ésima diferencia del proceso Z_t .

Series de tiempo no estacionarias en varianza

Un proceso que es estacionario en la media no es necesariamente estacionario en la varianza y en covarianza. Sin embargo, un proceso que no es estacionario en la media también será

no estacionario en varianza y en autocovarianza. Wei(1994)[26]

Aunque la diferenciación puede ser usada para reducir una serie no estacionaria homogénea a una serie estacionaria, muchas series de tiempo no son homogéneas. La no estacionaridad de estas series no se debe a que su función de media depende del tiempo, sino porque su función de varianzas y covarianzas depende del tiempo. Para reducir estas series a estacionaridad se necesitan transformaciones diferentes a la diferenciación.

Para estacionarizar una serie que no sea estacionaria en varianza frecuentemente se emplea una transformación potencial la cual puede estabilizar su varianza.

2.2. Familia de transformaciones de Box y Cox

2.2.1. Transformación de Box y Cox en la Regresión Lineal

El análisis de regresión lineal clásico se basa en los supuestos de que el término de error es aditivo, sigue una distribución normal y tiene varianza constante. Cuando estas hipótesis son seriamente violadas frecuentemente se diseña un nuevo modelo que tenga las características importantes del modelo original y satisfaga todos los supuestos por medio de la aplicación de una transformación adecuada a los datos o filtrando algunos datos que parecen sospechosos de ser atípicos. Es frecuente que el camino elegido por muchos investigadores sea la transformación de Box y Cox (1964)[5] la cual es utilizada con el objetivo de que los supuestos de aditividad, normalidad y varianza constante sean satisfechos aproximadamente.

La transformación de Box y Cox (1964)[5] trata de estimar el parámetro λ de una transformación de potencia sobre la variable dependiente del modelo de regresión lineal

$$y_i^{(\lambda)} = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} + \epsilon_i \quad (2-11)$$

donde

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(y_i) & \text{si } \lambda = 0 \end{cases} \quad (2-12)$$

es la familia de transformaciones de potencia de Box y Cox.

La función densidad de probabilidad de las observaciones no transformadas, y por consiguiente la función de verosimilitud en relación con las observaciones originales, se obtiene

multiplicando la función de densidad normal por el Jacobiano de la transformación. La transformación estimada se obtiene por medio de la maximización de la verosimilitud normal.

$$L(\lambda, \beta, \sigma^2 | y, X) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}(y(\lambda) - X\beta)'(y(\lambda) - X\beta)\right] J(\lambda, y) \quad (2-13)$$

donde

$$J(\lambda, y) = \prod_{i=1}^n y_i^{\lambda-1} \quad (2-14)$$

Aunque la transformación estimada posee las propiedades de los estimadores máximo verosímiles, no es robusta a la presencia de observaciones atípicas en la variable dependiente.

2.2.2. Transformación de Box y Cox en Series de Tiempo

Es muy común que en un proceso no estacionario su varianza cambie a medida que cambia su nivel, así:

$$Var(Z_t) = cf(\mu_t) \quad (2-15)$$

Para una constante c positiva y función f monótona.

En estos casos es posible encontrar una transformación $T(Z_t)$ de forma tal que $Var(T(Z_t))$ sea constante. Wei (1994)[26]

2.3. Observaciones atípicas

La literatura sobre las observaciones atípicas es muy amplia, a través de la historia son muchos los autores que se han dedicado a estudiar estas observaciones desde diferentes ámbitos, a categorizarlas, identificarlas y desarrollar tratamientos adecuados para mitigar los posibles impactos generados en la inferencia..

Las series de tiempo pueden contener observaciones anormales causados por errores o alteraciones inusuales que hacen parte del resto de las observaciones de la serie. McQuarrie y

Tsai (2003)[19]. Las series de tiempo se ven frecuentemente afectadas por este tipo de observaciones o eventos externos desconocidos, las observaciones generadas por estos eventos son llamadas observaciones atípicas (Outliers).

Se ha demostrado que la presencia de estas observaciones atípicas afecta la inferencia, generando poca confiabilidad o invalidez en los resultados, llevando a tomar decisiones inadecuadas sobre el verdadero proceso generador de los datos.

2.3.1. Observaciones atípicas (outliers) en series de tiempo

En el análisis de series de tiempo se ha estudiado la presencia de observaciones atípicas y se han propuesto procedimientos que permitan su detección y remoción de sus efectos. Fox(1972)[13] propone para modelos autoregresivos (AR) dos tipos de modelos de detección de outliers, Chang, Tiao y Chen (1988)[10] desarrollan y extienden la identificación y detección de outliers para modelos autorregresivos y de medias móviles (ARMA).

3. Metodología propuesta

En el análisis de series de tiempo generalmente es necesario transformar la varianza de la serie, para esto es común utilizar la familia de transformaciones introducida por Box y Cox en 1964[5], pero esta transformación puede verse afectada cuando existe presencia de observaciones atípicas en la serie.

Para el caso en estudio se pretende buscar que la transformación de la serie de tiempo Z_t , tenga en cuenta el efecto de la presencia de observaciones atípicas, obteniendo una estimación adecuada y robusta del parámetro λ de la familia de transformaciones de Box y Cox, mediante la extensión de la propuesta de Castaño (2012)[8] para las series de tiempo.

3.1. Transformación de Box y Cox

Para mostrar la adecuada funcionalidad de la transformación de Box y Cox cuando no existe presencia de observaciones atípicas, se simuló una serie con 200 observaciones usando un modelo autorregresivo y de medias móviles ARMA(1,1), donde $\lambda=0.5$, $\theta_0=10$, $\phi_1=0.4$, $\theta_1=0.8$ y $a_t \sim N(0, 1)$, ver gráfico 3-1.

Al aplicar la transformación de Box y Cox para esta simulación tenemos los resultados reflejados en el gráfico 3-2, donde se demuestra que λ se ejecuta de forma adecuada, ya que el empleo de dicha transformación proporciona $\hat{\lambda}=0.5$ sugiriendo $\sqrt{Z_t}$. Siendo coherente con los parámetros dados inicialmente, se verifica para este caso que la transformación introducida por Box y Cox (1964)[5] trabaja de forma apropiada cuando no existe presencia de observaciones atípicas.

Figura 3-1.: Modelo sin presencia de observaciones atípicas

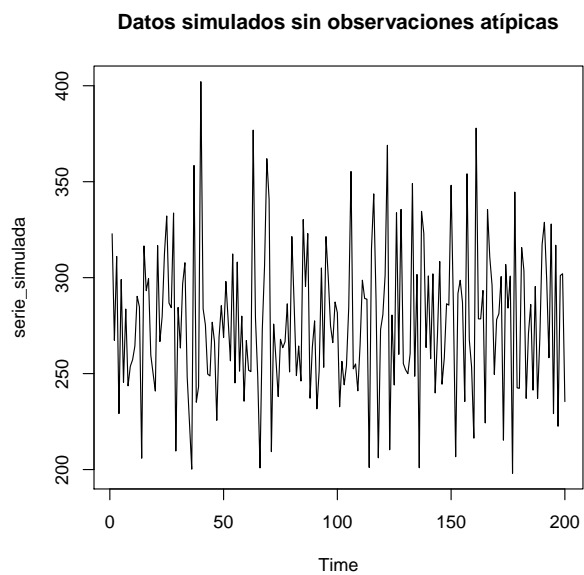
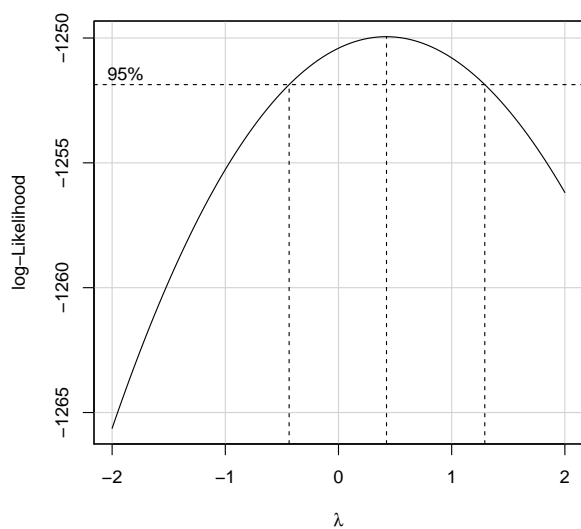


Figura 3-2.: Parámetro λ cuando no hay presencia de observaciones atípicas



3.2. Transformación de Box y Cox ante la presencia de observaciones atípicas

En el gráfico 3-3 se presenta una serie simulada con 200 observaciones usando un modelo autorregresivo y de medias móviles ARMA(1,1), donde $\lambda=0.5$, $\theta_0=10$, $\phi_1=0.4$, $\theta_1=0.7$ y $a_t \sim N(0,1)$ y una contaminación de 3 observaciones procedentes de una $N(0,25)$, correspondiente a observaciones atípicas aditivas. Al aplicar la transformación de Box y Cox para esta simulación tenemos los resultados reflejados en el gráfico 3-4, donde vemos que la aplicación de la transformación proporciona $\hat{\lambda}=-1$ lo cual sugiere una transformación no adecuada. La estimación del modelo produce $\hat{\theta}_0=0.0024$, $\hat{\phi}_1=0.3360$ y $\hat{\theta}_1=0.7371$, los resultados como verificación son interesantes para ilustrar que la simulación está acorde con el hecho demostrado de que la transformación no opera bien en presencia de los atípicos introducidos en la serie y en consecuencia los estimadores del modelo.

Sin embargo, si transformamos el modelo con el verdadero valor de $\lambda = 0,5$; y ajustamos el modelo, los resultados producen $\hat{\theta}_0=11.06713$, $\hat{\phi}_1=0.3353$ y $\hat{\theta}_1=0.7431$, donde se ve mejoría en los estimadores, pero siguen siendo sensibles a la presencia de observaciones atípicas.

Figura 3-3.: Modelo con presencia de observaciones atípicas

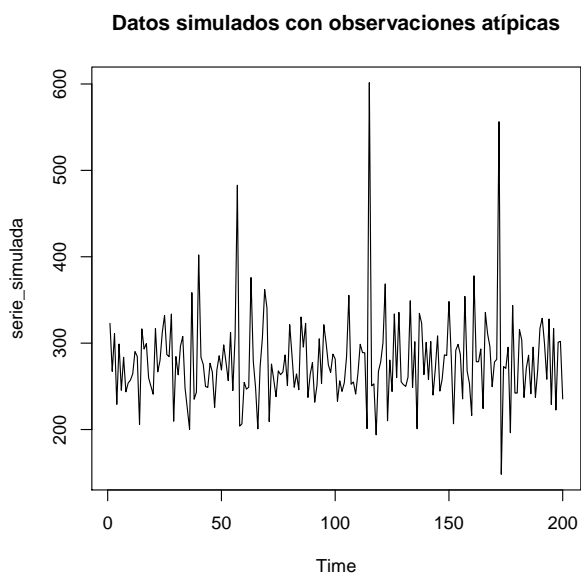
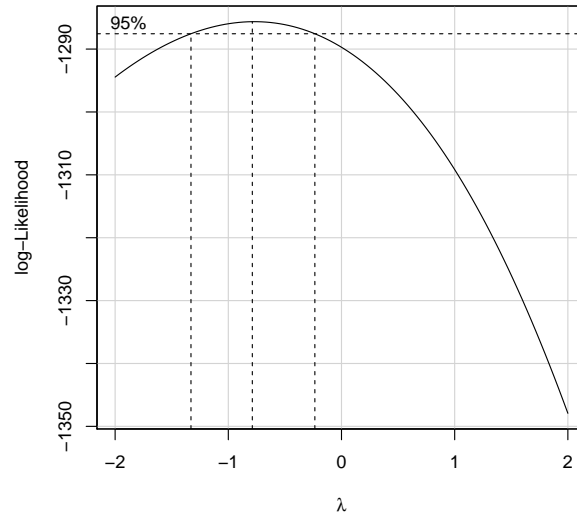


Figura 3-4.: Parámetro λ cuando hay presencia de observaciones atípicas

Se replicó el procedimiento anterior para un modelo de medias móviles MA de orden 1 MA(1) con parámetros $\theta_1=0.8$, $\mu=10$ y $a_t \sim N(0, 1)$, para un modelo autoregresivo de orden 1 AR(1) con parámetros $\theta_0=10$, $\phi_1=0.4$ y $a_t \sim N(0, 1)$, y para un modelo autorregresivo y de medias móviles ARMA de orden 1,1 ARMA(1,1) con parámetros $\theta_0=10$, $\theta_1=0.8$, $\phi_1=0.4$ y $a_t \sim N(0, 1)$, para valores de $\lambda=0.25, 0.5, 1$ se aplicó la transformación de Box y Cox. Posteriormente las series fueron contaminadas con 3 observaciones procedentes de una $N(0, 25)$ correspondiente a observaciones atípicas aditivas, se aplicó de nuevo la transformación y se realizó una comparación de los resultados que pueden observarse en las tablas a continuación.

Tabla 3-1.: Resultados datos simulados con $\lambda = 0,25$

$\lambda = 0,25$		
Modelo	sin contaminar	contaminado
AR(1)	0.2438	-0.4926
MA(1)	0.2472	0.2066
ARMA(1,1)	0.255	-0.3244

Tabla 3-2.: Resultados datos simulados con $\lambda = 0,5$

$\lambda = 0,5$		
Modelo	sin contaminar	contaminado
AR(1)	0.5084	-0.9728
MA(1)	0.5096	0.421
ARMA(1,1)	0.4936	-0.6248

Tabla 3-3.: Resultados datos simulados con $\lambda = 1$

$\lambda = 1$		
Modelo	sin contaminar	contaminado
AR(1)	0.9028	-1.7704
MA(1)	0.9686	0.8384
ARMA(1,1)	0.9814	-1.2248

Para las realizaciones se puede resaltar que la transformación de Box y Cox se comporta de forma adecuada cuando no hay existencia de observaciones atípicas, sin embargo, sucede lo contrario cuando se desea aplicar la transformación en presencia de observaciones atípicas, ya que el resultado arrojado no es el adecuado; es decir, los resultados de estas simulaciones son consistentes con los hallazgos de los autores que ya demostraron que la presencia de las observaciones atípicas afecta seriamente la estimación del parámetro de transformación λ y en consecuencia las estimaciones de los parámetros del modelo, siendo coherente con lo mencionado en las secciones anteriores.

3.3. Procedimiento propuesto

El procedimiento que se propone en este trabajo es una extensión de la propuesta de Castaño (2012)[8] donde se presenta un procedimiento alternativo no paramétrico y robusto que permite obtener la transformación de potencia en la familia de transformaciones de Box y Cox cuando existen observaciones atípicas en la variable dependiente en regresión lineal, el objetivo es extender la propuesta para series de tiempo.

El procedimiento que se propone trata de estimar el parámetro de transformación λ de tal manera que en el modelo (3-1)

$$Z_t^{(\lambda)} = \theta_0 + \phi_1 Z_{t-1}^{(\lambda)} + \phi_2 Z_{t-2}^{(\lambda)} + \dots + \phi_p Z_{t-p}^{(\lambda)} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3-1)$$

a_t sea aditivo, homocedástico y con distribución simétrica.

El proceso propuesto está definido en cinco etapas, con las cuales se pretende estimar un valor para el parámetro de transformación λ . Las etapas son las siguientes:

- i. Se propone aproximar a la serie Z_t usando un proceso autorregresivo de orden alto, aplicando el resultado obtenido por Said y Dickey (1984)[24] donde propone que el orden alto para el modelo autorregresivo p^* sea calculado así $p^* = n^{1/3}$, donde n es el tamaño de la serie, dado que en la práctica se desconoce el orden del modelo a estudiar,

$$Z_t = \theta_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_{p^*} Z_{t-p^*} + a_t \quad (3-2)$$

- ii. Definir un conjunto de valores para λ , generalmente el valor de λ se encuentra en el intervalo $[-2, 2]$.
- iii. Para cada valor de λ elegido, estime el modelo aproximado Z_t^λ y calcule los residuales $e_i(\lambda)$.
- iv. Obtenga los residuales normalizados como e_i^N , donde N significa normalizados

$$e_i^N(\lambda) = \frac{e_i(\lambda)}{MAD(e_i(\lambda))} \quad \text{donde} \quad MAD = \text{mediana}\{| e_i(\lambda) - \text{mediana}\{e_i(\lambda)\} |\}.$$

Este proceso elimina las diferentes unidades de medida en la función objetivo, introducidas al ir cambiando el valor de λ entre los valores $[-2, 2]$.

- v. Calcule los percentiles $\varepsilon_p(\lambda)$ y $\varepsilon_{1-p}(\lambda)$ de $e_i^N(\lambda)$ para varios valores de p , $0 < p < 1$. Obtenga:

$$\varepsilon_{0,5} - \frac{\varepsilon_p(\lambda) + \varepsilon_{1-p}(\lambda)}{2} \quad (3-3)$$

defina la función:

$$SA(\lambda) = \sum_p \left| \varepsilon_{0,5} - \frac{\varepsilon_p(\lambda) + \varepsilon_{1-p}(\lambda)}{2} \right| \quad (3-4)$$

Bajo el supuesto de que la transformación λ simetriza la distribución de errores

$$\varepsilon_{0,5} - \frac{\varepsilon_p(\lambda) + \varepsilon_{1-p}(\lambda)}{2} = 0 \quad (3-5)$$

para todo p , $0 < p < 1$

Por lo tanto, el valor $\hat{\lambda}$ que minimiza a $SA(\lambda)$ es la transformación de Box y Cox que simetriza la distribución de los errores, en otras palabras $\hat{\lambda}$ es el parámetro de transformación buscado.

3.4. Error estándar de la transformación estimada

Para el cálculo del error estándar se emplea la técnica del Bootstrap, Efron y Tibshirani (1986)[11].

- i. Calcule la transformación $\hat{\lambda}$ y obtenga los residuales del modelo ajustado.

$$Z_t^{(\hat{\lambda})} = \theta_0 + \phi_1 Z_{t-1}^{(\hat{\lambda})} + \phi_2 Z_{t-2}^{(\hat{\lambda})} + \dots + \phi_p Z_{t-p}^{(\hat{\lambda})} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3-6)$$

- ii. Obtenga una muestra aleatoria de tamaño m usando remuestreo con reemplazo de los residuales a_t . Sean a_t^* los residuales obtenidos. Construya los pseudo datos para la variable dependiente Z_t así

$$Z_t^* = (\theta_0 + \phi_1 Z_{t-1}^{(\hat{\lambda})} + \phi_2 Z_{t-2}^{(\hat{\lambda})} + \dots + \phi_p Z_{t-p}^{(\hat{\lambda})} + a_t^* - \theta_1 a_{t-1}^* - \theta_2 a_{t-2}^* - \dots - \theta_q a_{t-q}^*)^{(1/\hat{\lambda})} \quad (3-7)$$

- iii. Use el procedimiento propuesto para estimar λ en el modelo de los pseudo datos

$$Z_t^{*(\lambda)} = \theta_0 + \phi_1 Z_{t-1}^{(\lambda)} + \phi_2 Z_{t-2}^{(\lambda)} + \dots + \phi_p Z_{t-p}^{(\lambda)} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3-8)$$

Regrese a ii. y repita el proceso B veces. Sea $\hat{\lambda}_j^*$ el estimador de λ obtenido en la iteración $j = 1, 2, \dots, B$.

Obtenga la desviación estándar de $\hat{\lambda}$ usando su distribución bootstrap, es decir el error estándar de $\hat{\lambda}$ es

$$se(\hat{\lambda}) = \left[\frac{1}{B-1} \sum_{j=1}^B (\hat{\lambda}_j^* - \bar{\lambda}^*)^2 \right]^{0,5} \quad (3-9)$$

donde $\bar{\lambda}^* = \frac{1}{B} \sum \hat{\lambda}_j^*$

4. Experimento vía simulación

4.1. Procedimiento propuesto condicional

Para la realización del experimento se tendrá en consideración series simuladas con y sin datos atípicos, con $n=200,500,1000$ observaciones, para la estimación del parámetro λ . Luego de ejecutar el experimento con diferentes repeticiones se determinó que el número de repeticiones adecuado fuera de 3.000, debido a que los resultados se tornan estables.

En el experimento se generaron observaciones para el modelo

$$Z_t^{(\lambda)} = \theta_0 + \phi_1 Z_{t-1}^{(\lambda)} - \theta_1 a_{t-1} + a_t \quad (4-1)$$

donde $\phi_1=0.4$, $\theta_1=0.8$, $\theta_0=10$, $a_t \sim RB$, con valores de $\lambda=0.25,0.5,1$, adicionalmente se generaron observaciones para el mismo modelo con una observación atípica en el percentil 0.25, 0.5 y 0.75 con magnitudes de 8 y 10 desviaciones estándar de la media de los a_t , correspondientes a observaciones atípicas aditivas. Para la aplicación del método propuesto se requiere conocer el orden de la serie, sin embargo, en la práctica este se desconoce, debido a esto, cada serie se aproximó a un modelo autorregresivo de orden alto dependiendo el tamaño de n seleccionado $n=200,500,1000$, con $p^*=6,8,10$ respectivamente según la aproximación de Said y Dickey (1984)[24] donde propone que $p^* = n^{1/3}$.

Los resultados reportados en las tablas 4-1:4:18 están conformados por la media de las estimaciones, la estimación de la raíz cuadrada del error cuadrático medio (RECM) que está dado por $\sqrt{\sum_{s=1}^{3000} (\lambda_s - \lambda)^2 / 3000}$ y la estimación del sesgo promedio (SESGO) que está dado por $\sum_{s=1}^{3000} (\lambda_s - \lambda) / 3000$, donde λ_s es una estimación de λ de las 3000 repeticiones realizadas.

Como el objetivo es comparar los resultados de las diversas estimaciones de λ , evaluaremos las diferencias del RECM, así en cuanto mas pequeño sea el estimador de RECM y del SESGO mejor sera el proceso de estimación.

La función objetivo fue minimizada usando los percentiles para $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.80$ y 0.90 . Los cálculos se realizaron usando el paquete `car` y `FitAR`.

Las tablas 4-1:4-18 y gráficos 4-1:4-9 presentan los resultados de la estimación de λ por medio de transformación tradicional de Box y Cox (denominada Box-Cox en las tablas) y usando el método propuesto (denominada Propuesta condicional en las tablas).

Los resultados de los experimentos realizados muestran que cuando existe presencia de observaciones atípicas, de forma general, el método propuesto es más preciso y produce menos error que el método de Box y Cox tradicional. Además se observa que a medida que el tamaño muestral crece, el error cuadrático y sesgo decrece y el estimador propuesto converge al parámetro desconocido, mostrando la propiedad de consistencia del nuevo estimador para el parámetro λ .

4.2. Procedimiento propuesto incondicional

Para la realización de este trabajo se consideró también estudiar la transformación propuesta incondicional y comparar los resultados obtenidos con la transformación condicional.

En la sección anterior se estudió la transformación propuesta condicional en la cual se tiene en cuenta el orden del modelo, definido para este trabajo por la aproximación de Said y Dickey (1984)[24] donde propone que $p^* = n^{1/3}$. Se pretende comparar los resultados de condicionar la estimación al orden del modelo con la aplicación directa a la serie sin considerar orden alguno, para determinar si existe o no una mejor estimación del parámetro de transformación.

En el experimento se generaron observaciones para el modelo

$$Z_t^{(\lambda)} = \theta_0 + \phi_1 Z_{t-1}^{(\lambda)} - \theta_1 a_{t-1} + a_t \quad (4-2)$$

donde $\phi_1=0.4$, $\theta_1=0.8$, $\theta_0=10$, $a_t \sim RB$, con valores de $\lambda=0.25,0.5,1$, adicionalmente se generaron observaciones para el mismo modelo con una observación atípica en el percentil 0.25, 0.5 y 0.75 con magnitudes de 8 y 10 desviaciones estándar, correspondiente a observaciones atípicas aditivas, y tamaños de $n=200,500,1000$.

Los resultados reportados a continuación están conformados por la media de las estimaciones, la estimación de la raíz cuadrada del error cuadrático medio (RECM) que está dado por $\sqrt{\sum_{s=1}^{3000} (\lambda_s - \lambda)^2 / 3000}$ y la estimación del sesgo promedio (SESGO) que está dado por $\sum_{s=1}^{3000} (\lambda_s - \lambda) / 3000$ donde λ_s es una estimación de λ de las 3000 repeticiones realizadas.

Como el objetivo es comparar los resultados de las diversas estimaciones de λ , evaluaremos las diferencias del RECM, así en cuanto más pequeño sea el estimador de RECM y del SES-GO mejor será el proceso de estimación.

Las siguientes tablas 4-1:4-18 y gráficos 4-1:4-9 presentan los resultados de la estimación de λ por medio de transformación tradicional de Box y Cox (denominada Box-Cox en las tablas), el método propuesto (denominada Propuesta condicional en las tablas) y usando la transformación directa a la serie (denominada Propuesta incondicional en las tablas) .

Tabla 4-1.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,25 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.24683	-0.23960	0.24960	-0.07880	0.24690	0.02820
Propuesta condicional	0.24987	0.21927	0.25000	0.23460	0.24173	0.23547
Propuesta incondicional	0.24167	0.24817	0.25880	0.26090	0.24443	0.24550
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.21709	0.51125	0.14152	0.34541	0.10013	0.23652
Propuesta condicional	0.57719	0.50993	0.35400	0.33638	0.23994	0.23577
Propuesta incondicional	0.45449	0.44616	0.29048	0.28787	0.20106	0.19973
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.00317	-0.48960	-0.00040	-0.32880	-0.00310	-0.22180
Propuesta condicional	-0.00013	-0.03073	0.00000	-0.01540	-0.00827	-0.01453
Propuesta incondicional	-0.00833	-0.00183	0.00880	0.01090	-0.00557	-0.00450

Tabla 4-2.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,25 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.24443	-0.27687	0.24867	-0.14803	0.24957	-0.04170
Propuesta condicional	0.24357	0.21740	0.26413	0.24740	0.24790	0.23623
Propuesta incondicional	0.24840	0.25957	0.25617	0.25900	0.25060	0.25380
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.22319	0.54692	0.13658	0.41071	0.10010	0.30193
Propuesta condicional	0.57040	0.47623	0.35373	0.32052	0.24324	0.23549
Propuesta incondicional	0.45275	0.43951	0.27986	0.27793	0.20564	0.20394
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.00557	-0.52687	-0.00133	-0.39803	-0.00043	-0.29170
Propuesta condicional	-0.00643	-0.03260	0.01413	-0.00260	-0.00210	-0.01377
Propuesta incondicional	-0.00160	0.00957	0.00617	0.00900	0.00060	0.00380

Tabla 4-3.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,50 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.24443	-0.24453	0.24837	-0.07773	0.24737	0.02640
Propuesta condicional	0.24380	0.21083	0.24810	0.22910	0.25047	0.24600
Propuesta incondicional	0.24513	0.25020	0.24933	0.25147	0.25037	0.25240
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.21647	0.51802	0.13723	0.34340	0.10053	0.23730
Propuesta condicional	0.57356	0.49796	0.34408	0.32902	0.24856	0.24231
Propuesta incondicional	0.44626	0.44055	0.28002	0.27852	0.20126	0.20138
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.00557	-0.49453	-0.00163	-0.32773	-0.00263	-0.22360
Propuesta condicional	-0.00620	-0.03917	-0.00190	-0.02090	0.00047	-0.00400
Propuesta incondicional	-0.00487	0.00020	-0.00067	0.00147	0.00037	0.00240

Tabla 4-4.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,50 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.24583	-0.27607	0.24630	-0.14973	0.25173	-0.04133
Propuesta condicional	0.24477	0.22550	0.25510	0.24167	0.24720	0.24170
Propuesta incondicional	0.23503	0.24740	0.25797	0.26290	0.25327	0.25767
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.22157	0.54723	0.14010	0.41223	0.10221	0.30141
Propuesta condicional	0.57578	0.47288	0.35031	0.32423	0.24432	0.23556
Propuesta incondicional	0.44761	0.43380	0.29746	0.29387	0.20609	0.20609
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.00417	-0.52607	-0.00370	-0.39973	0.00173	-0.29133
Propuesta condicional	-0.00523	-0.02450	0.00510	-0.00833	-0.00280	-0.00830
Propuesta incondicional	-0.01497	-0.00260	0.00797	0.01290	0.00327	0.00767

Tabla 4-5.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,75 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.23637	-0.25083	0.24520	-0.08027	0.25053	0.02947
Propuesta condicional	0.23027	0.19763	0.24497	0.22897	0.25577	0.24607
Propuesta incondicional	0.24353	0.25037	0.24987	0.25000	0.25257	0.25410
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.22283	0.52271	0.13830	0.34617	0.10080	0.23415
Propuesta condicional	0.55474	0.49722	0.34993	0.32857	0.24461	0.23961
Propuesta incondicional	0.45091	0.44345	0.28258	0.27688	0.20757	0.20612
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.01363	-0.50083	-0.00480	-0.33027	0.00053	-0.22053
Propuesta condicional	-0.01973	-0.05237	-0.00503	-0.02103	0.00577	-0.00393
Propuesta incondicional	-0.00647	0.00037	-0.00013	0.00000	0.00257	0.00410

Tabla 4-6.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0,75 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.24703	-0.27663	0.24803	-0.14617	0.25010	-0.04187
Propuesta condicional	0.24937	0.22293	0.25003	0.23360	0.23980	0.23067
Propuesta incondicional	0.24873	0.26020	0.25117	0.25493	0.24720	0.24987
Raíz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	0.21915	0.54665	0.14002	0.40831	0.10020	0.30217
Propuesta condicional	0.57483	0.47524	0.34400	0.31906	0.24579	0.23845
Propuesta incondicional	0.45284	0.44351	0.28197	0.28081	0.20572	0.20557
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin outliers	Outliers
Box-Cox	-0.00297	-0.52663	-0.00197	-0.39617	0.00010	-0.29187
Propuesta condicional	-0.00063	-0.02707	0.00003	-0.01640	-0.01020	-0.01933
Propuesta incondicional	-0.00127	0.01020	0.00117	0.00493	-0.00280	-0.00013

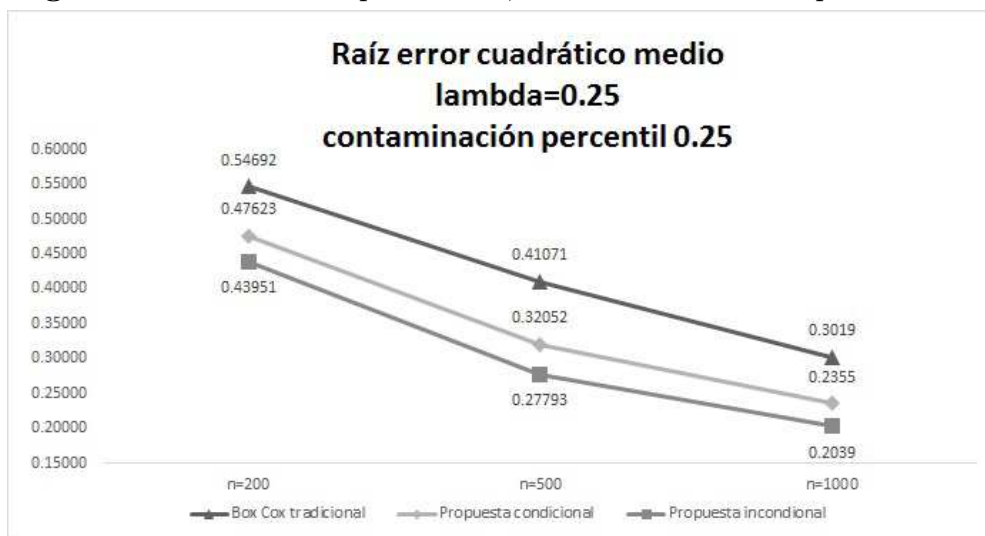
Figura 4-1.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0.25

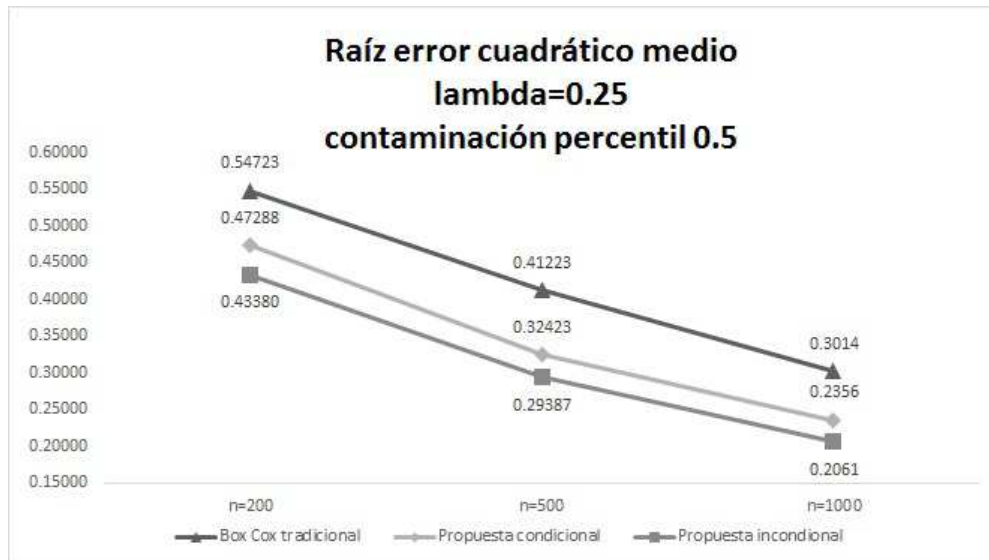
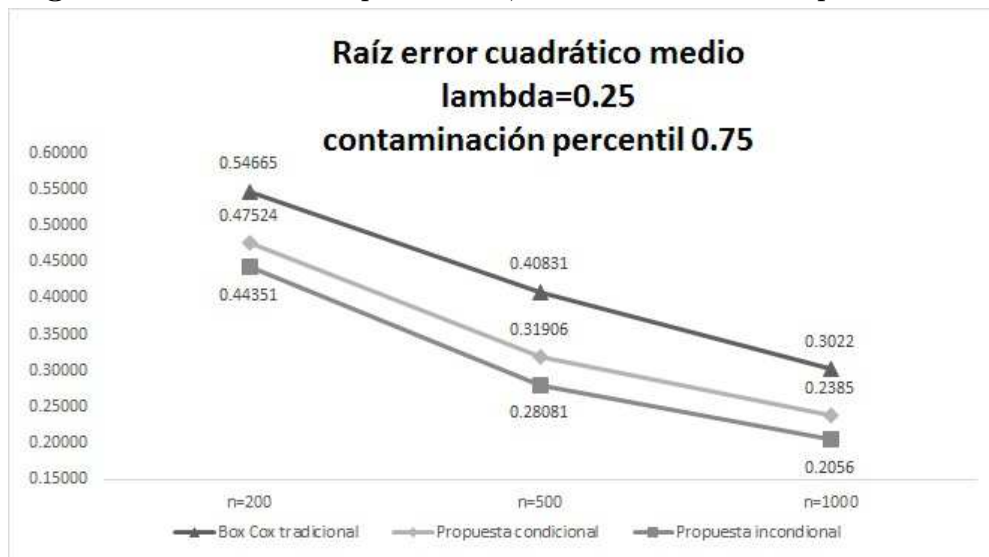
Figura 4-2.: Resultados para $\lambda = 0,25$ contaminada en el percentil 0.5**Figura 4-3.:** Resultados para $\lambda = 0,25$ contaminada en el percentil 0.75

Tabla 4-7.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,25 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.49827	-0.48823	0.49620	-0.16083	0.49370	0.05127
Propuesta condicional	0.42557	0.37673	0.49820	0.47177	0.50770	0.48973
Propuesta incondicional	0.47537	0.48960	0.50567	0.51140	0.49623	0.49890
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.43884	1.03229	0.27494	0.69032	0.19314	0.47347
Propuesta condicional	1.01490	0.93992	0.69131	0.66784	0.48513	0.47843
Propuesta incondicional	0.86075	0.84610	0.55284	0.55102	0.39722	0.39571
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.00173	-0.98823	-0.00380	-0.66083	-0.00630	-0.44873
Propuesta condicional	-0.07443	-0.12327	-0.00180	-0.02823	0.00770	-0.01027
Propuesta incondicional	-0.02463	-0.01040	0.00567	0.01140	-0.00377	-0.00110

Tabla 4-8.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,25 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.48743	-0.54377	0.50337	-0.28837	0.49843	-0.08577
Propuesta condicional	0.44747	0.41423	0.49213	0.47163	0.50147	0.48130
Propuesta incondicional	0.48453	0.51307	0.49793	0.51290	0.50997	0.51507
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.44263	1.08603	0.27199	0.81101	0.19845	0.60386
Propuesta condicional	1.01225	0.89034	0.68059	0.63297	0.46764	0.45684
Propuesta incondicional	0.85512	0.83699	0.54900	0.54215	0.45684	0.39616
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.01257	-1.04377	0.003373	-0.78837	-0.00157	-0.58577
Propuesta condicional	-0.05253	-0.08577	-0.00787	-0.02837	0.00147	-0.01870
Propuesta incondicional	-0.01547	0.01307	-0.00207	0.01290	0.00997	0.01507

Tabla 4-9.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,5 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.48247	-0.48363	0.50623	-0.15860	0.49717	0.05853
Propuesta condicional	0.43087	0.40063	0.51227	0.47860	0.49423	0.47743
Propuesta incondicional	0.46197	0.47133	0.51467	0.51740	0.50020	0.50160
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.44612	1.02983	0.27359	0.68840	0.19224	0.46730
Propuesta condicional	0.99237	0.91805	0.65757	0.63837	0.47256	0.47566
Propuesta incondicional	0.85725	0.84453	0.55286	0.54547	0.39571	0.39241
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.01753	-0.98363	0.00623	-0.65860	-0.00283	-0.44147
Propuesta condicional	-0.06913	-0.09937	0.01227	-0.02140	-0.00577	-0.02257
Propuesta incondicional	-0.03803	-0.02867	0.01467	0.01740	0.00020	0.00160

Tabla 4-10.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,5 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.49920	-0.54700	0.48547	-0.29817	0.49467	-0.08860
Propuesta condicional	0.43913	0.40410	0.50000	0.47527	0.50393	0.48673
Propuesta incondicional	0.46693	0.50237	0.48790	0.49767	0.49697	0.50047
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.44316	1.08573	0.27868	0.82238	0.19408	0.60651
Propuesta condicional	1.01055	0.88638	0.67869	0.63606	0.47514	0.46635
Propuesta incondicional	0.86218	0.84308	0.55668	0.55415	0.39140	0.39117
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.00080	-1.04700	-0.01453	-0.79817	-0.00533	-0.58860
Propuesta condicional	-0.06087	-0.09590	0.00000	-0.02473	0.00393	-0.01327
Propuesta incondicional	-0.03307	0.00237	-0.01210	-0.00233	-0.00303	0.00047

Tabla 4-11.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,75 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.49933	-0.48927	0.50000	-0.15627	0.49380	0.05663
Propuesta condicional	0.44953	0.41470	0.48580	0.45190	0.49933	0.47960
Propuesta incondicional	0.49733	0.50703	0.50317	0.50913	0.48887	0.49027
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.44073	1.03283	0.27122	0.685307	0.19425	0.46978
Propuesta condicional	1.00620	0.92629	0.67209	0.65891	0.47917	0.46883
Propuesta incondicional	0.83668	0.82519	0.56032	0.55801	0.39700	0.39504
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.00067	-0.98927	0.00000	-0.65627	-0.00620	-0.44337
Propuesta condicional	-0.05047	-0.08530	-0.01420	-0.04810	-0.00067	-0.02040
Propuesta incondicional	-0.00267	0.00703	0.00317	0.00913	-0.01113	-0.00973

Tabla 4-12.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0,75 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.47733	-0.55113	0.49483	-0.29260	0.49203	-0.08840
Propuesta condicional	0.42243	0.41513	0.50053	0.46417	0.48657	0.47513
Propuesta incondicional	0.46860	0.50117	0.51593	0.52420	0.48137	0.48833
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.43132	1.09035	0.27848	0.81597	0.20031	0.60630
Propuesta condicional	0.99676	0.87941	0.70697	0.66140	0.467793	0.45881
Propuesta incondicional	0.84747	0.83030	0.55827	0.55128	0.40134	0.40060
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.02267	-1.05113	-0.00517	-0.79260	-0.00797	-0.58840
Propuesta condicional	-0.07757	-0.08487	0.00053	-0.03583	-0.013437	-0.02487
Propuesta incondicional	-0.03140	0.00117	0.01593	0.02420	-0.01863	-0.01167

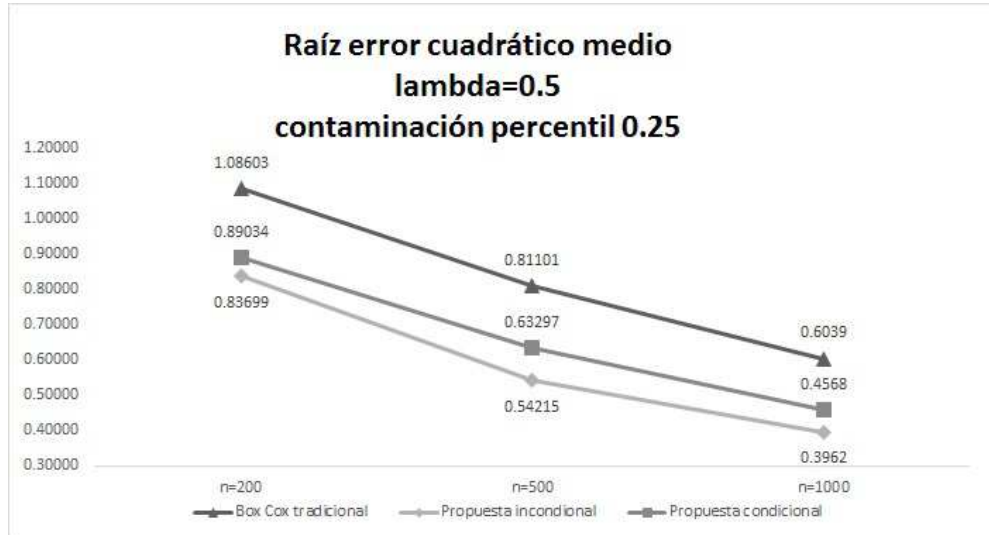
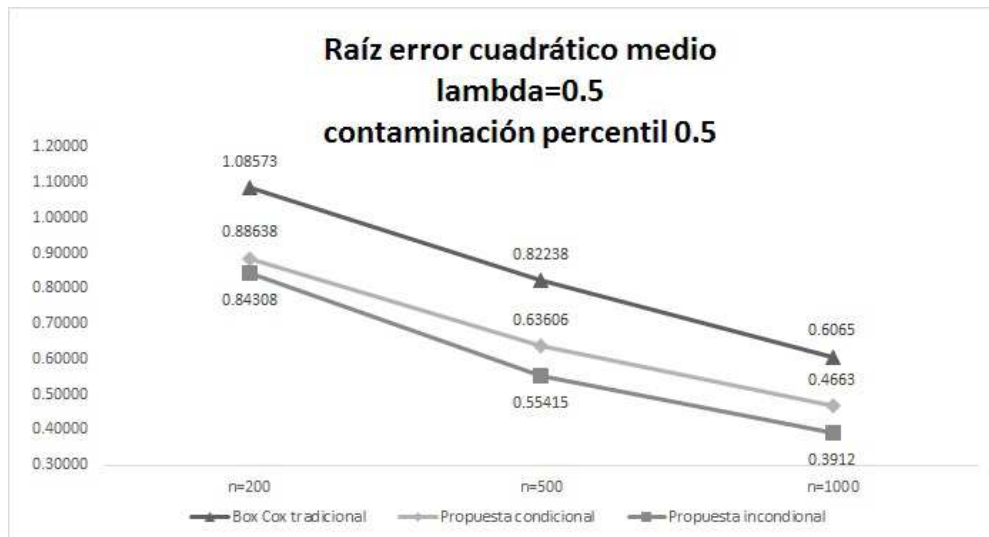
Figura 4-4.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0.25Figura 4-5.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0.5

Figura 4-6.: Resultados para $\lambda = 0,5$ contaminada en el percentil 0.75

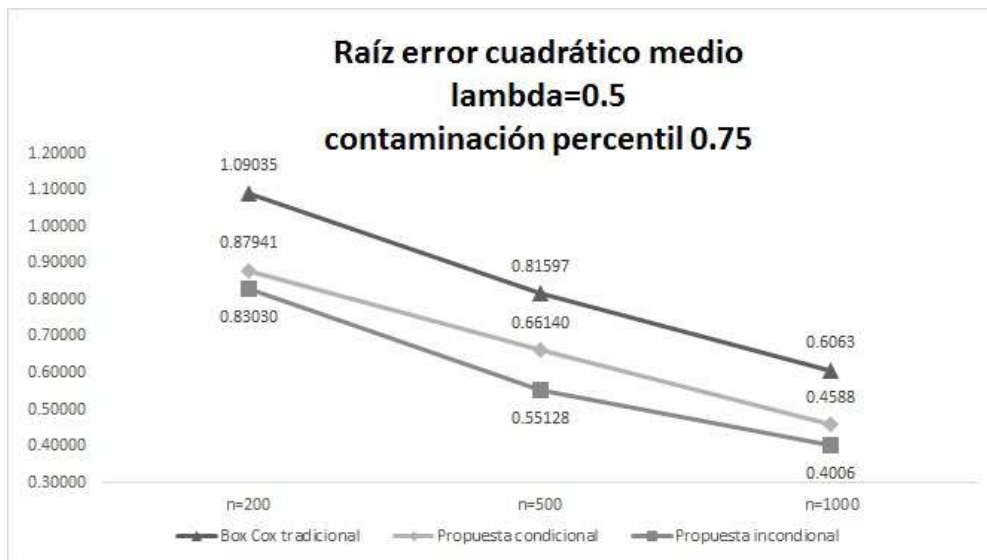


Tabla 4-13.: Resultados para $\lambda = 1$ contaminada en el percentil 0,25 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.93900	-0.97237	0.97423	-0.33060	1.00247	0.12643
Propuesta condicional	0.56600	0.50720	0.77880	0.75090	0.92473	0.90090
Propuesta incondicional	0.73183	0.76963	0.87350	0.88390	0.96887	0.97343
Raíz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.77505	2.05031	0.54661	1.39091	0.38198	0.92418
Propuesta condicional	1.39185	1.32760	1.07222	1.04830	0.80911	0.79261
Propuesta incondicional	1.26734	1.24673	0.95773	0.94695	0.72346	0.72010
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.06100	-1.97237	-0.02577	-1.33060	0.00247	-0.87357
Propuesta condicional	-0.43400	-0.49280	-0.22120	-0.24910	-0.07527	-0.09910
Propuesta incondicional	-0.26817	-0.23037	-0.12650	-0.11610	-0.03113	-0.02657

Tabla 4-14.: Resultados para $\lambda = 1$ contaminada en el percentil 0,25 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.94343	-1.08893	0.97300	-0.59190	1.00303	-0.17283
Propuesta condicional	0.57853	0.55160	0.78810	0.76153	0.92857	0.90127
Propuesta incondicional	0.75457	0.78873	0.89523	0.91280	0.98197	0.98960
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.78425	2.15862	0.53878	1.63821	0.38751	1.20571
Propuesta condicional	1.39006	1.28832	1.06046	1.04492	0.81662	0.81604
Propuesta incondicional	1.28965	1.25747	0.95409	0.94215	0.70311	0.69666
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.05657	-2.08893	-0.02700	-1.59190	0.00303	-1.17283
Propuesta condicional	-0.42147	-0.44840	-0.21190	-0.23847	-0.07143	-0.09873
Propuesta incondicional	-0.24543	-0.21127	-0.10477	-0.08720	-0.01803	-0.01040

Tabla 4-15.: Resultados para $\lambda = 1$ contaminada en el percentil 0,5 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.92757	-0.96090	0.99690	-0.31423	0.98997	0.11193
Propuesta condicional	0.54080	0.44970	0.79060	0.75643	0.90337	0.87963
Propuesta incondicional	0.67840	0.70507	0.87640	0.89230	0.96940	0.97177
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.78253	2.04028	0.52928	1.37356	0.38304	0.93733
Propuesta condicional	1.40752	1.36566	1.05094	1.05193	0.82843	0.83503
Propuesta incondicional	1.31419	1.28827	0.95167	0.93636	0.72248	0.71783
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.07243	-1.96090	-0.00310	-1.31423	-0.01003	-0.01003
Propuesta condicional	-0.45920	-0.55030	-0.18973	-0.24357	-0.09663	-0.12037
Propuesta incondicional	-0.32160	-0.29493	-0.12360	-0.10770	-0.03060	-0.02823

Tabla 4-16.: Resultados para $\lambda = 1$ contaminada en el percentil 0,5 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.89507	-1.09837	0.99090	-0.57380	0.99027	-0.17227
Propuesta condicional	0.49693	0.51443	0.79060	0.77717	0.91720	0.88960
Propuesta incondicional	0.65030	0.69150	0.88673	0.90937	0.96217	0.97380
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.80223	2.17084	0.53596	1.62119	0.38255	1.20429
Propuesta condicional	1.40833	1.29401	1.06809	1.03169	0.84201	0.83605
Propuesta incondicional	1.32591	1.29966	0.94951	0.93866	0.71593	0.71340
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.10493	-2.09837	-0.00910	-1.57380	-0.00973	-1.17227
Propuesta condicional	-0.50307	-0.48557	-0.20940	-0.22283	-0.08280	-0.11040
Propuesta incondicional	-0.34970	-0.30850	-0.11327	-0.09063	-0.03783	-0.02620

Tabla 4-17.: Resultados para $\lambda = 1$ contaminada en el percentil 0,75 magnitud 8σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.91753	-0.96363	0.98530	-0.31257	1.00507	0.12357
Propuesta condicional	0.51057	0.46793	0.79093	0.75697	0.93380	0.90333
Propuesta incondicional	0.68257	0.71117	0.86560	0.87320	0.98567	0.98997
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.79038	2.04350	0.52004	1.37117	0.37694	0.92655
Propuesta condicional	1.41886	1.35990	1.06269	1.05881	0.80916	0.81457
Propuesta incondicional	1.31723	1.29301	0.94331	0.93495	0.70785	0.70544
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.08247	-1.96363	-0.01470	-1.31256	0.00507	-0.87643
Propuesta condicional	-0.48943	-0.53207	-0.20907	-0.24303	-0.06620	-0.09667
Propuesta incondicional	-0.31743	-0.28883	-0.13440	-0.12680	-0.01433	-0.01003

Tabla 4-18.: Resultados para $\lambda = 1$ contaminada en el percentil 0,75 magnitud 10σ

Media de las estimaciones	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.93497	-1.08137	0.97873	-0.58013	0.99110	-0.17580
Propuesta condicional	0.55287	0.53950	0.76107	0.74810	0.89907	0.88280
Propuesta incondicional	0.76797	0.81860	0.89780	0.91990	0.96407	0.97450
Raiz cuadrada del error cuadrático	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	0.78177	2.15326	0.53579	1.62634	0.39322	1.20853
Propuesta condicional	1.39703	1.27371	1.07424	1.03857	0.84184	0.82057
Propuesta incondicional	1.25837	1.22169	0.94230	0.92717	0.71790	0.70985
Sesgo promedio	N=200	N=200	N=500	N=500	N=1000	N=1000
	Sin outliers	Outliers	Sin outliers	Outliers	Sin Outliers	Outliers
Box-Cox	-0.06503	-2.08137	-0.02127	-1.58013	-0.00890	-1.17580
Propuesta condicional	-0.44713	-0.46050	-0.23893	-0.25190	-0.10093	-0.11720
Propuesta incondicional	-0.23203	-0.18140	-0.10220	-0.08010	-0.03593	-0.02550

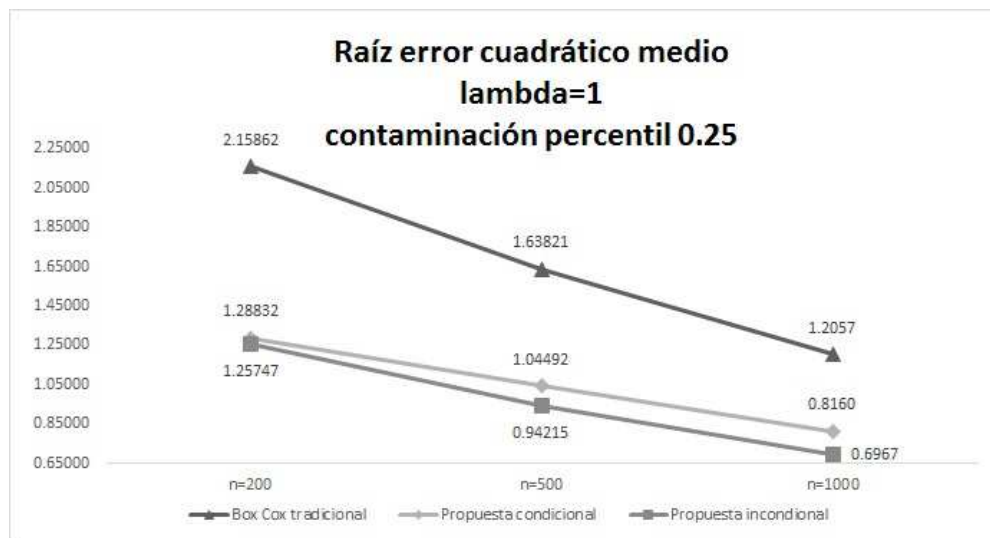
Figura 4-7.: Resultados para $\lambda = 1$ contaminada en el percentil 0.25

Figura 4-8.: Resultados para $\lambda = 1$ contaminada en el percentil 0.5

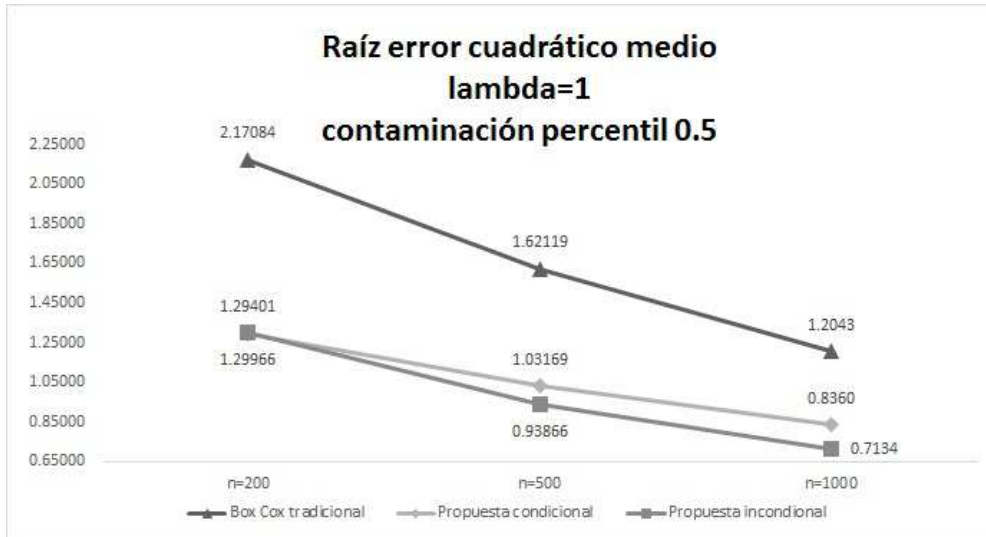
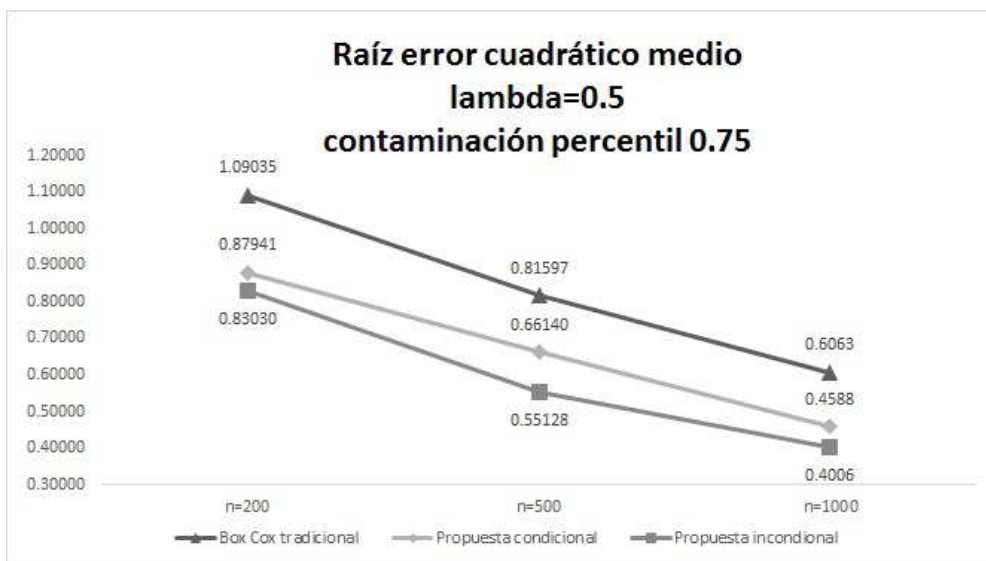


Figura 4-9.: Resultados para $\lambda = 1$ contaminada en el percentil 0.75



Los resultados de los experimentos realizados muestran que cuando existe presencia de observaciones atípicas aditivas, de forma general, el método propuesto condicional e incondicional son más precisos y producen menos error que el método de Box y Cox tradicional. Además se observa que a medida que el tamaño muestral crece, el error cuadrático y sesgo decrece y el estimador propuesto converge al parámetro desconocido, mostrando la propiedad de consistencia del nuevo estimador para el parámetro λ .

5. Análisis de resultados

Con el fin de cumplir los objetivos planteados en este trabajo, se analizó una variedad de casos para determinar los diferentes impactos que la presencia de observaciones atípicas aditivas podría generar al momento de estimar el parámetro de transformación.

En el primer análisis realizado en la sección 3.1, se observa que al aplicar la transformación en el caso donde no hay existencia de observaciones atípicas aditivas, la estimación del parámetro λ arrojado es coherente con el valor simulado, mostrando un desempeño correcto.

En el segundo análisis realizado en la sección 3.2, se observa que la transformación tradicional de Box y Cox se distorsiona y es poco precisa ante la presencia de observaciones atípicas aditivas, mostrando que la presencia de dichas observaciones afecta la estimación del verdadero valor del parámetro λ y en consecuencia, los estimadores del modelo. Es importante recordar que en la práctica es muy común encontrarnos con este tipo de escenarios, llevándonos a buscar metodologías alternativas para estacionarizar la varianza.

En las tablas 4-1:4-18, reportadas en la sección 4, se presentan los resultados de la estimación de λ por medio de la transformación tradicional de Box y Cox (denominado Box Cox en las tablas) y usando el método propuesto (denominado Propuesta condicional en las tablas), donde se muestra que ante la presencia de observaciones atípicas aditivas la transformación propuesta parece ser más precisa y produce menos error que el método de Box y Cox tradicional, se observa también que la magnitud de la observación atípica afecta la estimación del parámetro de transformación. Además en las gráficas de la sección 4 se observa para cada valor de λ analizado, que a medida que el tamaño muestral crece, el error cuadrático y sesgo decrece y el estimador propuesto converge al parámetro desconocido, mostrando la propiedad de consistencia del nuevo estimador para el parámetro λ .

En las tablas 4-1:4-18, reportadas en la sección 4 se presentan los resultados al estudiar la transformación propuesta condicional e incondicional cuando existe presencia de observaciones atípicas aditivas, donde se verifica que el desempeño de la transformación propuesta incondicional es más adecuado y preciso.

Para los experimentos realizados en la sección 4, los resultados muestran que cuando existe presencia de observaciones atípicas aditivas es más preciso el procedimiento propuesto

que el método tradicional de Box y Cox, sin embargo, se aprecia que la transformación propuesta parece no ser precisa en muestras pequeñas, pero si en muestras moderadas y grandes.

6. Conclusiones y recomendaciones

6.1. Conclusiones

Se confirma que la estimación del parámetro λ de la transformación de Box y Cox se ve afectada por causa de la presencia de observaciones atípicas aditivas; al existir presencia de dichas observaciones en la muestra, se afecta seriamente la estimación del parámetro y en consecuencia la estimación de los parámetros del modelo, arrojando resultados alejados del modelo real, haciendo que en la práctica para este tipo de casos sea más conveniente no transformar la serie y asumir las consecuencias de estudiar un modelo no estacionario en varianza.

La estimación robusta y no paramétrica para el parámetro λ de la transformación de Box y Cox propuesto en este trabajo presenta un buen desempeño en los casos analizados para diferentes valores de λ . En la mayoría de los casos analizados la transformación propuesta obtuvo mejor desempeño que la transformación tradicional cuando hay presencia de observaciones atípicas aditivas.

La transformación propuesta incondicional muestra mejor desempeño que aplicar la transformación propuesta condicional, esto demuestra que al aplicar la transformación directamente a la serie la estimación del parámetro de transformación es más adecuado y preciso.

Por otro lado, la estimación del parámetro λ de la transformación de Box y Cox tradicional, muestra ser más adecuada que la transformación de Box y Cox propuesta para los casos donde no hay presencia de observaciones atípicas aditivas.

Los resultados muestran que sin importar la ubicación de la observación atípica esta afectará la estimación del parámetro λ .

Se considera que la transformación propuesta es de gran utilidad para el análisis de series de tiempo, ya que es flexible en el sentido que logra disminuir el efecto de las observaciones atípicas aditivas sobre la estimación del parámetro de transformación λ , escenario que muy frecuentemente resulta imposible de evadir y este tipo de metodologías permite dar otras soluciones para el tratamiento de series de tiempo no estacionarias en varianza.

6.2. Recomendaciones

Con el fin de realizar buenos diagnósticos en la práctica, es conveniente continuar investigando alternativas para trabajar con series de tiempo no estacionarias en varianza, buscando que la estimación del parámetro de transformación λ tenga en cuenta el efecto de la presencia de observaciones atípicas.

Se considera valioso estudiar el impacto de otro tipo de observaciones atípicas que podrían presentarse en el análisis de series de tiempo y poder determinar si el tipo de observación atípica afecta de igual forma la estimación del parámetro de transformación λ .

A. Algoritmos y simulaciones

- Simulación de las series, con existencia o no de observaciones atípicas

Descripción:

Genera n valores de un modelo autorregresivo y de medias móviles.

Argumentos:

n número de observaciones a simular

ϕ_1 Parámetro autorregresivo

θ_1 Parámetro medias móviles

θ_0 Nivel de la serie

λ_{sim} Potencia

Resultado:

Serie de tiempo con existencia o no de observaciones atípicas

Código:

```
#####  
###FUNCION PARA SIMULAR LAS SERIES CON Y SIN CONTAMINACION#####  
#####  
  
simulacion<-function(n,phi1,theta1,theta0,lambda_sim)  
{  
  at=rnorm(n, mean = 0, sd =1)  
  at_cont=at  
  at_cont[n*0.25+30]=mean(at)+10*sd(at)  
  at_cont[n*0.50+30]=mean(at)+10*sd(at)  
  at_cont[n*0.75+30]=mean(at)+10*sd(at)  
  
  ##simulacion serie sin contaminar  
  Z_ARMA11=rep(0, times=n)  
  Z_ARMA11[1]=(theta0+phi1*(theta0/(1-phi1))lambda_sim+at[1]-
```

```

theta1*mean(at))^(1/lambda_sim)
for(t in 2:n)
{
Z_ARMA11[t]=(theta0+phi1*Z_ARMA11[t-1]^lambda_sim+at[t]-
theta1*at[t-1])^(1/lambda_sim)
}
# eliminacion de las primeras 30 observaciones
nobs=c(1:n)
serie=subset(Z_ARMA11, nobs>30)

##simulacion serie conntaminada
Z_ARMA11_cont=rep(0, times=n)
Z_ARMA11_cont[1]=(theta0+phi1*(theta0/(1-phi1))^lambda_sim+at_cont[1]-
theta1*mean(at_cont))^(1/lambda_sim)
for(t in 2:n)
{
Z_ARMA11_cont[t]=(theta0+phi1*Z_ARMA11_cont[t-1]^lambda_sim+at_cont[t]-
theta1*at_cont[t-1])^(1/lambda_sim)
}
# eliminacion de las primeras 30 observaciones
nobs=c(1:n)
serie_cont=subset(Z_ARMA11_cont, nobs>30)
list(Serie=serie,Contaminada=serie_cont)

}

```

- Función de la transformación propuesta, requiere la serie a transformar y el orden

Descripción:

Función para estimar el parámetro de transformación λ robusto y no paramétrico propuesto en este trabajo.

Argumentos:

P Orden autorregresivo alto por el cual se va aproximar, $p^* = T^{1/3}$ propuesto por Said y Dickey (1984)[24]

Serie serie a transformar

Resultado:

El parámetro de transformación λ propuesto en este trabajo

Codigo:

```
#####
#####FUNCION TRANSFORMACION PROPUESTA#####
#####

transformacion_propuesta<-function(serie,P)
{
lambda= seq(-2, 2, 1/10)
#-----
#Transformo cada serie
#-----
serie_trans=NULL
for(y in 1:length(lambda))
{
serie_trans<-cbind(serie_trans,serie^lambda[y]-1/lambda[y])
}
serie_trans[,21]<-log(serie)
serie_trans[,31]<-serie
serie_trans
#-----
#-----
#regresion del modelo cara cada lambda
#Para cada serie extraer los residuales
residualesar_alto=NULL
residualesar_alto_norm=NULL
for(m in 1:length(lambda))
{
#####
#residuales ar alto###
#####
residualesar_alto<-cbind(residualesar_alto,FitAR(serie_trans[,m],P)$res)
residualesar_alto_norm<-cbind(residualesar_alto_norm,
residualesar_alto[,m]/(median(abs(residualesar_alto[,m])-
median(residualesar_alto[,m]))))
}
prob1=0.1
prob2=0.2
prob3=0.3
```

```
prob4=0.4
prob5=0.5
prob6=0.6
prob7=0.7
prob8=0.8
prob9=0.9

prob2<-c(prob1,prob2,prob3,prob4,prob5,prob6,prob7,prob8,prob9)

SA_ar_alto=NULL
for(t in 1:length(lambda))
{
SA_ar_alto[t]=0
ABS_ar_alto=0
for(u in 1:9)
{
ABS_ar_alto=sum(abs(quantile(residualesar_alto_norm[,t],prob2[5])-
(((quantile(residualesar_alto_norm[,t],prob2[u])+
quantile(residualesar_alto_norm[,t],1-prob2[u]))/2))))
SA_ar_alto[t]=SA_ar_alto[t]+ABS_ar_alto
}
resultado_ar_alto<-cbind(lambda,SA_ar_alto)
resu_ar_alto<-resultado_ar_alto[which.min(resultado_ar_alto[,2])]
}
propuesta_ar_alto<-resultado_ar_alto[which.min(resultado_ar_alto[,2])]
matrix(c(propuesta_ar_alto), nrow =1, ncol = 1)
}
```

Bibliografía

- [1] Anthony Atkinson and Marco Riani. *Robust diagnostic regression analysis*. Springer, 2000.
- [2] Anthony Curtis Atkinson and Neil Shephard. Deletion diagnostics for transformations of time series. *Journal of forecasting*, 15(1):1–17, 1996.
- [3] Richard J Beckman and R Dennis Cook. Outlier.....s. *Technometrics*, 25(2):119–149, 1983.
- [4] Peter J Bickel and Kjell A Doksum. An analysis of transformations revisited. *Journal of the american statistical association*, 76(374):296–311, 1981.
- [5] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.
- [6] Moshe Buchinsky. Quantile regression, box-cox transformation model, and the us wage structure, 1963–1987. *Journal of Econometrics*, 65(1):109–154, 1995.
- [7] Raymond J Carroll and David Ruppert. Robust estimation in heteroscedastic linear models. *The annals of statistics*, pages 429–441, 1982.
- [8] Elkin Castaño. Una estimación no paramétrica y robusta de la transformación box-cox para el modelo de regresión. *Lecturas de Economía*, 75(75):89–106, 2012.
- [9] Gary Chamberlain. Quantile regression, censoring, and the structure of wages. In *Advances in Econometrics: Sixth World Congress*, volume 2, pages 171–209, 1994.
- [10] Ih Chang, George C Tiao, and Chung Chen. Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204, 1988.
- [11] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*, volume 57. CRC press, 1994.
- [12] Bernd Fitzenberger, Ralf A Wilke, and Xuan Zhang. *A note on implementing box-cox quantile regression*. Zentrum für Europ. Wirtschaftsforschung, 2004.
- [13] Anthony J Fox. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 350–363, 1972.

-
- [14] Clive WJ Granger and Paul Newbold. Forecasting transformed series. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 189–203, 1976.
- [15] Victor M Guerrero. Time-series analysis supported by power transformations. *Journal of forecasting*, 12(1):37–48, 1993.
- [16] Peter J Lenk and Chih-Ling Tsai. Transformations and dynamic linear models. *Journal of Forecasting*, 9(3):219–232, 1990.
- [17] Agustin Maravall. An application of nonlinear time series forecasting. *Journal of Business & Economic Statistics*, 1(1):66–74, 1983.
- [18] Alfio Marazzi and Victor J Yohai. Robust box–cox transformations based on minimum residual autocorrelation. *Computational statistics & data analysis*, 50(10):2752–2768, 2006.
- [19] Allan D McQuarrie and Chih-Ling Tsai. Outlier detections in autoregressive models. *Journal of Computational and Graphical Statistics*, 12(2), 2003.
- [20] Alan Pankratz. *Forecasting with dynamic regression models*, volume 935. John Wiley & Sons, 2012.
- [21] James L Powell. Estimation of monotonic regression models under quantile restrictions. *Nonparametric and semiparametric methods in Econometrics*, (Cambridge University Press, New York, NY), pages 357–384, 1991.
- [22] Marco Riani. Extensions of the forward search to time series. *Studies in Nonlinear Dynamics & Econometrics*, 8(2), 2004.
- [23] Marco Riani. Robust transformations in univariate and multivariate time series. *Econometric Reviews*, 28(1-3):262–278, 2008.
- [24] Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.
- [25] Chih-Ling Tsai and Xizhi Wu. Transformation-model diagnostics. *Technometrics*, 34(2):197–202, 1992.
- [26] William Wu-Shyong Wei. *Time series analysis*. Addison-Wesley publ, 1994.