



COLEGIO DE POSTGRADUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

**MODELOS DE REGRESIÓN ROBUSTA  
CON APLICACIÓN EN ECONOMETRÍA**

ALDO RODRÍGUEZ CORTÉS

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2019

---

---

**CARTA DE CONSENTIMIENTO DE USO DE LOS DERECHOS DE AUTOR Y DE LAS REGALIAS COMERCIALES DE PRODUCTOS DE INVESTIGACION**

En adición al beneficio ético, moral y académico que he obtenido durante mis estudios en el Colegio de Postgraduados, el que suscribe ALDO RODRÍGUEZ CORTÉS, Alumno (a) de esta Institución, estoy de acuerdo en ser participe de las regalías económicas y/o académicas, de procedencia nacional e internacional, que se deriven del trabajo de investigación que realicé en esta institución, bajo la dirección del Profesor DR. PAULINO PÉREZ RODRÍGUEZ, por lo que otorgo los derechos de autor de mi tesis

MODELOS DE REGRESIÓN ROBUSTA CON APLICACIÓN EN ECONOMETRÍA

y de los productos de dicha investigación al Colegio de Postgraduados. Las patentes y secretos industriales que se puedan derivar serán registrados a nombre del colegio de Postgraduados y las regalías económicas que se deriven serán distribuidas entre la Institución, El Consejero o Director de Tesis y el que suscribe, de acuerdo a las negociaciones entre las tres partes, por ello me comprometo a no realizar ninguna acción que dañe el proceso de explotación comercial de dichos productos a favor de esta Institución.

Montecillo, Mpio. de Texcoco, Edo. de México, a 02 de MAYO de 2019



Firma del  
Alumno (a)



DR. PAULINO PÉREZ RODRÍGUEZ

Vo. Bo. del Consejero o Director de Tesis

La presente tesis titulada: **Modelos de Regresión Robusta con Aplicación en Econometría**, realizada por el alumno: **Aldo Rodríguez Cortés**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## MAESTRO EN CIENCIAS

### SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

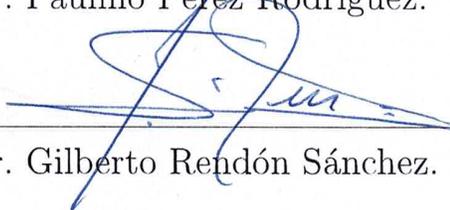
#### CONSEJO PARTICULAR

CONSEJERO



Dr. Paulino Pérez Rodríguez.

ASESOR



Dr. Gilberto Rendón Sánchez.

ASESOR



Dr. Mario Alberto Vázquez Peña.

# Modelos de Regresión Robusta con Aplicación en Econometría

Aldo Rodríguez Cortés, M. en C.

Colegio de Postgraduados, 2019

## RESUMEN

En econometría y finanzas, son de gran interés los modelos que determinan la relación cuantitativa entre el rendimiento en exceso para un valor determinado y el rendimiento en exceso en la cartera del mercado en que cotiza, sin embargo, una crítica constante es que, generalmente, se supone que los rendimientos son simétricos y normalmente distribuidos, no obstante, se ha encontrado que la distribución del rendimiento financiero es de cola pesada y ligeramente asimétrica.

Existe una amplia literatura estadística sobre distribuciones asimétricas, las cuales recientemente han recibido mucha atención en la investigación para resolver problemas como el descrito, para esto diferentes autores han propuesto varias distribuciones asimétricas, incluyendo versiones de la *Normal Asimétrica (SN)*, *t-Asimétrica (ST)*, *Cauchy Asimétrica (SC)*, *Laplace Asimétrica (SL)*, entre otras; la mayoría de las cuales se construyen mezclando dos distribuciones o escalando una distribución. En este trabajo se propone el uso de una clase de distribución asimétrica univariada, la *ST*, que incluye un parámetro de forma que le permite modelar la asimetría, representada aquí por dos parametrizaciones distintas, una de ellas construida aplicando un factor de escala aleatorio adecuado a una variable *SN*, replicando el mecanismo que genera la distribución habitual de *t-Student* y por lo que puede considerarse una generalización que incluye como caso particular la *SN*; la otra, generada por medio de una mezcla de distribuciones *Normales*.

En ambos casos, las representaciones estocásticas de una variable aleatoria *ST* permiten hacer inferencia Bayesiana mediante el uso de técnicas de Cadenas de Markov Monte Carlo (MCMC) para ajustar el modelo propuesto en el contexto de Regresión Lineal Múltiple (RLM). Se presentan los desarrollos teórico-matemáticos que permiten la obtención de la distribución posterior de los parámetros para la *ST*, utilizando la técnica de Aumentación de Datos. Se optó por hacer uso de la formulación actualmente más utilizada de la *ST* para la creación de un programa en el paquete estadístico *R* que permite realizar el ajuste, evaluar el poder predictivo y la bondad de ajuste del modelo. Se muestra el uso de la propuesta mediante estudios de simulación y la aplicación a un conjunto de datos reales de rendimientos financieros ampliamente estudiado. Finalmente, se presentan dos Pruebas de Bondad de Ajuste para la distribución *ST*, una basada en el coeficiente de Correlación de Pearson y otra basada en el cálculo del estadístico de Kolmogorov-Smirnov mediante *Bootstrap* paramétrico.

**Palabras clave:** Distribución *t-Asimétrica (ST)*, Inferencia Bayesiana, Aumentación de Datos, Pruebas de Bondad de Ajuste, Rendimientos Financieros.

# Robust Regression Models with Application in Econometrics

Aldo Rodríguez Cortés, M. en C.

Colegio de Postgraduados, 2019

## ABSTRACT

In econometrics and finance, models that determine the quantitative relationship between excess return for a given value and excess performance in the market in which it is quoted are of great interest, however, a constant criticism is that, generally, it is assumed that the returns are symmetric and normally distributed, however, it has been found that the distribution of financial return is thick-tailed and slightly skewed.

There is vast statistical literature on skewed distributions, which recently have received much attention in research to solve problems such as the one described, for this different authors have proposed several skewed distributions, including versions of the *Skew-Normal* ( $SN$ ), *Skew- $t$*  ( $ST$ ), *Skew-Cauchy* ( $SC$ ), *Skew-Laplace* ( $SL$ ), among others; most of which are built by mixing two distributions or scaling a distribution. In this work we propose the use of a univariate skewed distribution class, the  $ST$ , which includes a parameter that allows it to model the asymmetry, represented here by two different parameterizations, one of them constructed applying a random scale factor appropriate to a variable  $SN$ , replicating the mechanism that generates the usual *t-Student* distribution which can be considered a generalization that includes as particular case the  $SN$  distribution; the other distribution, generated by a mixture of *Normal* distributions.

In both cases, the stochastic representations of a random  $ST$  variable allow us to make Bayesian inference by using Markov Chain Monte Carlo (MCMC) techniques to fit the proposed model in the context of Multiple Linear Regression (MLR). The theoretical-mathematical developments that allow the obtaining of the posterior distribution of the parameters for the  $ST$  are presented, using the Data Augmentation algorithm. It was decided to make use of the currently more accredited formulation of the  $ST$  for the creation of a program in the  $R$  statistical package that allows to fit the model, to evaluate the predictive power and goodness of fit of the model. The use of the proposal is shown through simulation studies and the application to a set of real data of financial returns widely studied. Finally, two Goodness of Fit Tests for the  $ST$  distribution are presented, one based on the Pearson Correlation Coefficient and another based on the calculation of a *Bootstrap* version of the Kolmogorov-Smirnov Goodness of fit test.

**Keywords:** *Skew- $t$*  distribution ( $ST$ ), Bayesian Inference, Data Augmentation, Goodness of Fit Tests, Financial Returns.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico brindado durante la realización de mis estudios de postgrado.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Dr. Paulino Pérez Rodríguez, por su excelente dirección, paciencia, profesionalismo, dedicación, e interés mostrado en orientarme para poder llevar a buen término este trabajo de investigación. Por haberme permitido trabajar con usted, el tiempo y el amplio conocimiento compartido durante mis estudios de postgrado y, principalmente, por su amistad, muchas gracias.

Dr. Gilberto Rendón Sánchez y Dr. Mario Alberto Vázquez Peña, por el tiempo dedicado a la revisión de este escrito y las observaciones realizados.

A todos los profesores, por haberme transmitido su invaluable conocimiento y contribuir así a mi formación académica.

A Aracely Yazmín Pérez Rodríguez y Heriberto Hernández Ruíz, por su paciencia, gran apoyo, compañerismo, inigualable compañía y entrañable amistad, sin la cual me habría sido más difícil llegar a la meta. Para ustedes, mi mayor y más sincero agradecimiento.

A mis compañeros de clases, personal administrativo y todos aquellos que de alguna u otra manera fueron copartícipes de esta tarea.

## DEDICADO A:

*Mis padres:*

***F. Rafael Rodríguez Hernández y Marina R. Cortés Ibáñez.***

Por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo y el amor mantenido a través del tiempo.

*Mis hermanos:*

***Rafael y Nallely.***

Por abrirse paso en la vida y mostrarme el camino a seguir, de quienes he aprendido aciertos y de momentos difíciles. Los quiero hermanos.

*“No nos dejemos corromper por un escepticismo estéril y deprimente; no nos desalentemos ante la tristeza de ciertas horas que pasan sobre las naciones. Vivir en la serena paz de los laboratorios y las bibliotecas. Preguntémonos primero: ¿qué he hecho por instruirme? y, después, al ir progresando, ¿qué he hecho por mi patria? Hasta que llegue el día en que podamos sentir la íntima satisfacción de pensar en que de alguna manera hemos contribuido al progreso y bienestar de la humanidad.”*

***Louis Pasteur (1822-1895).***

# CONTENIDO

<b>Resumen</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>Lista de Tablas</b>	<b>xi</b>
<b>Lista de Figuras</b>	<b>xii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	3
1.2. Metodología . . . . .	3
1.3. Organización del Trabajo . . . . .	3
<b>2. Revisión de literatura</b>	<b>5</b>
2.1. Algoritmo de Aceptación y Rechazo . . . . .	5
2.2. Aumentación de datos . . . . .	6
2.3. Inferencia Bayesiana . . . . .	8
2.4. Muestreador de Gibbs . . . . .	9
2.5. Algoritmo de Metropolis . . . . .	11
2.6. Caso Caminata Aleatoria Metropolis (RWM) . . . . .	12

## CONTENIDO

---

2.7. Caso Muestreo Independiente (IS) . . . . .	13
2.8. Regresión con Errores aleatorios <i>Normales</i> . . . . .	13
2.9. Regresión con Errores aleatorios <i>t-Student</i> . . . . .	15
2.9.1. Distribución <i>t-Student</i> . . . . .	15
2.9.2. Regresión Bayesiana con Errores <i>t-Student</i> . . . . .	17
2.10. Regresión con Errores aleatorios <i>Normal-Asimétricos</i> . . . . .	21
2.10.1. Distribución <i>Normal-Asimétrica</i> . . . . .	21
2.10.2. Regresión Bayesiana con Errores <i>Normal-Asimétricos (BSN)</i> . . . . .	22
2.11. Distribución <i>t-Asimétrica (ST)</i> . . . . .	31
2.11.1. <i>ST</i> Azzalini y Capitanio (2003), y Gupta (2003) . . . . .	31
2.11.2. <i>ST</i> Zhu y Galbraith (2010) . . . . .	33
2.11.3. <i>ST</i> Fernández y Steel (1998) . . . . .	35
2.11.4. <i>ST</i> Jones y Faddy (2003) . . . . .	35
2.12. Elección de una adecuada Distribución <i>ST</i> . . . . .	40
<b>3. Regresión con Errores Aleatorios <i>ST</i></b> . . . . .	<b>41</b>
3.1. Regresión Bayesiana con Errores <i>ST</i> de Azzalini y Capitanio (2003), ( <i>BST</i> ) . . . . .	41
3.2. Regresión Bayesiana con Errores <i>ST</i> de Zhu y Galbraith (2010) . . . . .	52
3.3. Estudios de Simulación . . . . .	60
<b>4. Aplicación con datos reales</b> . . . . .	<b>64</b>
4.1. Conjunto de Datos de Rendimientos Financieros de la compañía Martin Marietta . . . . .	64
4.2. Ajuste del modelo <i>Normal</i> . . . . .	68

## CONTENIDO

---

4.3. Ajuste del modelo <i>BST</i> . . . . .	70
4.4. Prueba de Bondad de Ajuste de Correlación . . . . .	74
4.5. Prueba de Bondad de Ajuste usando <i>Bootstrap</i> Paramétrico . . . . .	78
<b>5. Conclusiones</b>	<b>80</b>
<b>Referencias</b>	<b>80</b>
<b>Anexos</b>	<b>85</b>
Anexo A: Código en <i>R</i> para ajustar el modelo de regresión Bayesiana con errores aleatorios <i>t-Student</i> . . . . .	85
Anexo B: Código en <i>R</i> para ajustar el modelo de regresión Bayesiana con errores aleatorios <i>SN</i> ( <i>BSN</i> ). . . . .	92
Anexo C: Código en <i>R</i> para ajustar el modelo de regresión con errores aleatorios <i>ST</i> de Azzalini y Capitanio (2003), ( <i>BST</i> ). . . . .	99
Anexo D: Código en <i>R</i> para el ajuste del modelo de regresión <i>Normal</i> , <i>t-Student</i> , <i>BSN</i> y <i>BST</i> para el conjunto de datos Martin Marietta. . . . .	108

# LISTA DE TABLAS

3.1. Comparación del ajuste de los modelos de regresión <i>BST</i> , <i>Normal</i> , <i>t-Student</i> , <i>BSN</i> y <i>ST</i> de la librería <b>SN</b> para un conjunto de datos simulados. . . . .	62
4.1. Datos: Rendimientos en Exceso, Compañía Martin Marietta (Butler <i>et al.</i> , 1990). . . . .	66
4.2. Estadísticas descriptivas. Conjunto de datos Martin Marietta. . . . .	67
4.3. Modelo clásico de regresión lineal. . . . .	68
4.4. Pruebas formales de Normalidad para los errores del MCRL. . . . .	70
4.5. Modelo de Regresión con errores <i>ST</i> . . . . .	71
4.6. Comparación del ajuste de los modelos de regresión <i>BST</i> , <i>Normal</i> , <i>t-Student</i> , <i>BSN</i> y <i>ST</i> de la librería <b>SN</b> para el conjunto de datos Martin Marietta. . . . .	73
4.7. Prueba de Bondad de Ajuste de Correlación para los residuales del modelo <i>BST</i> . . . . .	77
4.8. Prueba de Bondad de Ajuste para los residuales del modelo <i>BST</i> usando <i>Bootstrap</i> Paramétrico . . . . .	79

# LISTA DE FIGURAS

2.1. Gráfico de la densidad <i>t-Student</i> . . . . .	16
2.2. Gráfico de la densidad <i>Normal Asimétrica</i> . . . . .	23
2.3. Gráfico de la densidad <i>ST</i> definida por Azzalini y Capitanio (2003) y Gupta (2003). . . . .	37
2.4. Gráfico de la densidad <i>ST</i> definida por Zhu y Galbraith (2010). . . . .	38
2.5. Gráfico de la densidad <i>ST</i> definida por Fernández y Steel (1998). . . . .	39
3.1. Gráfico del funcionamiento del Algoritmo de Aceptación y Rechazo para el muestreo de $p(v_i resto)$ . . . . .	48
3.2. Histograma de residuales y gráfico de dispersión de los valores observados vs valores predichos del modelo <i>ST</i> para el conjunto de datos simulados. . . . .	62
3.3. Gráfico de las densidades posteriores de los parámetros estimados $\hat{\beta}_0$ , $\hat{\beta}_1$ , $\hat{\omega}$ y $\hat{\rho}$ del modelo <i>ST</i> para el conjunto de datos simulados. . . . .	63
4.1. Métodos gráficos para probar Normalidad en los errores del MCRL. . . . .	69
4.2. Gráfico de las densidades posteriores de los parámetros estimados $\hat{\beta}_0$ , $\hat{\beta}_1$ , $\hat{\omega}$ y $\hat{\rho}$ del modelo <i>ST</i> para el conjunto de datos Martin Marietta. . . . .	72
4.3. Histograma de residuales y gráfico de dispersión de los valores observados vs valores predichos del modelo <i>ST</i> para el conjunto de datos Martin Marietta. . . . .	73

# Capítulo 1

## Introducción

La econometría según [Hernández y Zúñiga \(2013\)](#), desde un enfoque moderno, se ha de considerar como la adecuada interacción entre teoría económica, matemática y estadística. De esta manera se pretende explicar fenómenos económicos mediante la definición de modelos matemáticos que son verificados por procedimientos estadísticos. Así se distingue entre la Teoría Econométrica y la Econometría Aplicada. La primera se refiere al conjunto de técnicas estadísticas que tienen por objeto la estimación y verificación de los modelos; y la segunda, al conjunto de aplicaciones de los métodos econométricos teóricos para alcanzar un conocimiento cuantitativo útil de los fenómenos económicos.

Los modelos econométricos se basan fundamentalmente en los modelos lineales y en el análisis de regresión como herramienta principal para la obtención de estimaciones con la finalidad de llegar a la toma de decisiones de índole económica o de política haciendo uso de aproximaciones frecuentistas, lo que se llama la corriente clásica, que domina la práctica econométrica. Por otra parte, los modelos econométricos fundados en la Teoría Bayesiana permiten hacer inferencia incorporando información inicial, a los datos muestrales, de modelos y parámetros en una forma coherente a través del Teorema de Bayes, con la finalidad de incorporar adecuadamente la incertidumbre en un problema de decisión ([Riascos, 2009](#)).

Los supuestos en que se basa el modelo clásico de regresión lineal (MCRL) son, entre otros, la linealidad en los parámetros, la normalidad en los errores, la homogeneidad de varianzas. Dichos supuestos restringen el estudio de los fenómenos económicos, y de cualquier tipo en general, debido a que el horizonte de predicción es limitado por la propia forma funcional de los modelos, la bondad de ajuste que presentan estos, y las propias variables causales utilizadas. Si un modelo es ajustado conforme al MCRL y no cumple estos supuestos, evidentemente se representarán incorrectamente los resultados, por ejemplo, los errores estándar de los coeficientes podrían estar sesgados, conduciendo a valores de las correspondientes estadísticas de prueba y *p-valores* incorrectos, con lo

## 1. Introducción

---

cual las pruebas de hipótesis llevarían a conclusiones equivocadas sobre la significancia de las covariables incluidas en el modelo. A su vez, si la estimación de los coeficientes es incorrecta, las predicciones serán distorsionadas y de poca utilidad práctica.

El uso de una clase limitada de distribuciones da como resultado una clase de inferencias limitada y potencialmente inapropiada. Muchos problemas quedan fuera del rango de los modelos convencionales, y los modelos deben elegirse para que se ajusten a la ciencia y los datos subyacentes, no simplemente por su conveniencia analítica o computacional. En situaciones como la presencia de valores atípicos, si, por conveniencia, se usan modelos simplistas, es importante preguntarse ¿de qué manera la inferencia depende de los valores extremos y de las suposiciones del modelo que no se pueden evaluar? Los modelos basados en la distribución *Normal* son notoriamente “no robustos” a los valores atípicos, en el sentido de que un solo dato aberrante puede afectar fuertemente la inferencia para todos los parámetros en el modelo, incluso aquellos con poca conexión directa a la observación atípica. En el marco Bayesiano, se puede reducir la influencia de observaciones aberrantes al reemplazar el modelo de población *Normal* por una familia de distribuciones de colas más pesadas, lo que permite la posibilidad de observaciones extremas. Por cola pesada, se entiende a una distribución con un contenido de probabilidad relativamente alto lejos del “centro” (Gelman *et al.*, 2014).

Un fenómeno de interés dentro del área econométrica y financiera que sirve para ejemplificar la problemática descrita anteriormente son los modelos que determinan la relación cuantitativa entre el rendimiento en exceso para una compañía en particular y el rendimiento en exceso en la cartera de cierto mercado de valores, en donde suele suponerse que dichos rendimientos siguen una distribución *Normal*, cuando en realidad, su distribución es de cola pesada y asimétrica.

Existe una tendencia general en la literatura estadística en dirección a métodos más flexibles para representar características de los datos en la medida de lo posible y reducir así supuestos poco realistas al momento de los análisis realizados (Russell-Noriega y González-Farías, 2002). Por ejemplo, existen diferentes clases paramétricas de distribuciones con algunas características particulares deseables, es el caso de la familia de distribuciones asimétricas, entre las que destacan, la *Normal Asimétrica (SN)*, *t-Asimétrica (ST)*, *Cauchy Asimétrica (SC)*, *Laplace Asimétrica (SL)*, entre otras.

Con la finalidad de proponer una herramienta estadística adecuada para el análisis de fenómenos similares al del área econométrica y financiera descrito, en este estudio se aborda la distribución *t-Asimétrica (ST)* univariada para ajustar modelos de Regresión Lineal Múltiple (RLM) en el contexto Bayesiano. Se utilizan dos parametrizaciones de la distribución, una de ellas construida aplicando un factor de escala aleatorio adecuado a una variable *SN*, replicando el mecanismo que genera la distribución habitual de *t-Student* y por lo que puede considerarse una generalización que incluye como caso particular la *SN*; la otra, generada por medio de una mezcla de distribuciones *Normales*.

### 1.1. Objetivos

**Generales:** Proponer modelos de regresión paramétrica con distribuciones de errores asimétricos.

**Particulares:** Aplicación de modelos propuestos a casos reales en ciencias económicas.

### 1.2. Metodología

Para el desarrollo de este trabajo se inicia con una revisión de la literatura acerca de los modelos más comunes utilizados en regresión robusta. Esto con la finalidad de conocer las herramientas actuales y disponibles con las que se está abordando este problema.

Se revisa el método de inferencia Bayesiana así como las herramientas que permiten su aplicación computacional, Cadenas de Markov Monte Carlo (Muestreador de Gibbs y Algoritmo de Metropolis), así mismo, los modelos de regresión con errores aleatorios *t-Student* y *Normal Asimétrico*.

Se analizan diferentes alternativas para la distribución *t-Asimétrica* univariada propuestas por distintos autores, las cuales cuentan con diferentes características propias y se elige una parametrización que resulta conveniente para el problema de estudio.

Se presentan los resultados obtenidos de los desarrollos matemáticos que permiten el ajuste de un modelo de regresión lineal múltiple con errores *ST*.

Los modelos presentados en este trabajo de investigación se ajustaron utilizando el software de uso libre *R* ([R Core Team, 2018](http://www.R-project.org/)), el cual puede ser descargado en el sitio <http://www.R-project.org/>.

### 1.3. Organización del Trabajo

En el capítulo 2 se realiza una breve revisión de la metodología para la Inferencia Bayesiana y se presentan algunas técnicas de Cadenas de Markov Monte Carlo (MCMC). Así mismo, se presenta el Modelo Clásico de Regresión Lineal (MCRL) y se plantean algunos problemas presentes en cierto tipo de análisis estadísticos dando paso a los modelos de regresión robusta que asumen diferentes distribuciones a la *Normal* en los errores; se analizan a detalle las propuestas que suponen una distribución *t-Student*

### 1.3. Organización del Trabajo

---

y *Normal Asimétrica (SN)*. Finalmente en este mismo capítulo se analizan diferentes alternativas para la distribución *t-Asimétrica* univariada (*ST*), la cual es el núcleo de la propuesta de este trabajo de investigación.

A continuación, en el capítulo 3, se presentan los desarrollos matemáticos para la obtención de las distribuciones condicionales de los parámetros de la distribución *ST* que permiten la implementación de un algoritmo de Gibbs y Metropolis para ajustar el modelo de regresión lineal múltiple (RLM) con errores *ST*. Al final de este capítulo se realizan algunos estudios de simulación en los que se comparan las diferentes propuestas presentadas anteriormente para ajustar el modelo de RLM con errores que siguen distribuciones distintas de la *Normal*.

Enseguida, en el capítulo 4, se realiza la aplicación del modelo de RLM con errores aleatorios *ST* propuesto a un conjunto de datos financieros reales ampliamente estudiado en distintos trabajos previos, con el fin de mostrar la utilidad de la propuesta desarrollada en esta investigación.

Posteriormente se encuentra el capítulo correspondiente a las Conclusiones (5), en donde se mencionan los principales puntos que pudieron ser extraídos del análisis de la propuesta desarrollada en este trabajo.

Finalmente se incluyen algunos Anexos que son, principalmente, los códigos creados en el paquete estadístico *R* (R Core Team, 2018), para el desarrollo de los análisis estadísticos presentados en esta investigación.

# Capítulo 2

## Revisión de literatura

En este capítulo se presenta un resumen de la teoría relacionada con los temas que serán utilizados en capítulos posteriores. En específico, los modelos de regresión robusta actualmente utilizados, es decir, regresión con errores aleatorios *t-Student* y *Normal Asimétricos (SN)*, y diferentes alternativas para la distribución *t-Asimétrica* univariada.

### 2.1. Algoritmo de Aceptación y Rechazo

Suponga que deseamos obtener muestras de una densidad  $p(\theta|\mathbf{y})$ , llamada densidad objetivo, o en su caso, una densidad no normalizada  $q(\theta|\mathbf{y})$  con  $p(\theta|\mathbf{y}) = \frac{q(\theta|\mathbf{y})}{\int q(\theta|\mathbf{y}) d\theta}$ .

El algoritmo de Aceptación y Rechazo es un método de Monte Carlo que permite realizar dicho trabajo y se utiliza, generalmente, para el muestreo de distribuciones univariadas y, a menudo, para generar distribuciones multivariadas truncadas, si la proporción de la masa de densidad en la parte truncada no es cercana a 1.

Este algoritmo requiere de una distribución llamada Generadora de Candidatos,  $g(\theta)$ , positiva definida para todo  $\theta$  para los cuales  $p(\theta|\mathbf{y}) > 0$  que tiene las siguientes propiedades:

- Es posible muestrear de la densidad de probabilidades proporcional a  $g(\cdot)$ . No se requiere que  $g(\cdot)$  integre 1, pero  $\int_{\mathbb{R}} g(\theta) d\theta < \infty$ .
- La razón de importancia  $\frac{p(\theta|\mathbf{y})}{g(\theta)}$  debe estar acotada superiormente, es decir, debe existir alguna constante  $M$  para la cual  $\frac{p(\theta|\mathbf{y})}{g(\theta)} \leq M$  para todo  $\theta$  ([Gelman et al.](#),

## 2.2. Aumentación de datos

---

2014).

Así, de acuerdo a [Tanner \(1996\)](#), el método de Aceptación y Rechazo es como se detalla en el Algoritmo 1.

<p><b>Algoritmo 1:</b> Algoritmo de Aceptación y Rechazo.</p> <p><b>input</b> : Para generar muestras de <math>p(\theta \mathbf{y})</math>, el algoritmo de Aceptación y Rechazo requiere la especificación previa de la Distribución Generadora de Candidatos, <math>g(\theta)</math>.</p> <p><b>output:</b> Muestra aleatoria de <math>p(\theta \mathbf{y})</math>.</p> <ol style="list-style-type: none"><li>1 Obtener <math>\theta^*</math> que resulta de maximizar la razón de importancia <math>m(\theta)</math>, <math display="block">\frac{p(\theta \mathbf{y})}{g(\theta)} = m(\theta).</math></li><li>2 Obtener la constante <math>M = m(\theta^*)</math></li><li>3 <b>for</b> <math>i</math> <b>in</b> <math>1:B</math> <b>do</b> /* con B número de muestras deseadas */</li><li>4     <b>reject</b>=TRUE. /* Rejection step */</li><li>5     <b>while</b> <i>reject</i> <b>do</b></li><li>6         Generar un candidato <math>\theta^c</math> de <math>g(\theta)</math>.</li><li>7         Generar <math>u</math>, <math>U \sim U(0, 1)</math>.</li><li>8         <b>if</b> <math>u \leq \frac{p(\theta^c \mathbf{y})}{Mg(\theta^c)}</math> <b>then</b></li><li>9             aceptar <math>\theta^c</math> como muestra de <math>p(\theta \mathbf{y})</math>.</li><li>10            <b>reject</b>=FALSE</li><li>11         <b>else</b></li><li>12             Rechazar el candidato <math>\theta^c</math></li><li>13         <b>end</b></li><li>14     <b>end</b></li><li>15 <b>end</b></li></ol>
--

El algoritmo de Aceptación y Rechazo permite muestrear indirectamente de  $p(\theta|\mathbf{y})$ , si se conoce su forma funcional y se proporciona una muestra aleatoria de una distribución aproximada  $g(\theta)$ . Se debe tener en cuenta que la probabilidad de aceptar un candidato  $\theta^c$  es igual a  $\int p(\theta|\mathbf{y}) d\theta/M$ . Por lo tanto, si la aproximación de  $g(\theta)$  a  $p(\theta|\mathbf{y})$  es pobre en algunas regiones (forzando a que la probabilidad anterior sea pequeña), entonces el algoritmo puede tener una baja tasa de aceptación.

## 2.2. Aumentación de datos

La idea de la aumentación de datos surge naturalmente en problemas de valores perdidos, como lo es en las formas estándar de completar las celdas faltantes en tablas

## 2.2. Aumentación de datos

---

de dos vías equilibradas. Así, la aumentación se refiere a un esquema de aumentar los datos observados para que sean más fácil de analizar. Este método es usado por el algoritmo EM (Esperanza-Maximización) para resolver problemas de estimación por Máxima Verosimilitud. En situaciones en las que la función de verosimilitud no tiene forma cerrada, las estimaciones de Máxima Verosimilitud y los errores estándar asociados no son confiables para hacer inferencias válidas. Desde el punto de vista Bayesiano, hay que calcular los parámetros de la distribución posteriores de interés. Si la aumentación de datos se puede usar en la estimación de Máxima Verosimilitud, entonces, de igual manera, se debería poder usar en el cálculo de la distribución posterior.

La idea básica es simple, los datos observados  $\mathbf{y}$  son aumentados por la cantidad  $\mathbf{z}$  (no observados), que se conoce como datos latentes. Se asume que si  $\mathbf{y}$  y  $\mathbf{z}$  son conocidos, entonces el problema es más sencillo de analizar, es decir, la densidad posterior de los datos aumentados  $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$  puede ser calculada. Pero la densidad posterior que deseamos es  $p(\boldsymbol{\theta}|\mathbf{y})$  lo que puede ser difícil de calcular directamente. Sin embargo, si uno puede generar múltiples valores de  $\mathbf{z}$  a partir de la distribución predictiva  $p(\mathbf{z}|\mathbf{y})$  (i.e. imputaciones múltiples de  $\mathbf{z}$ ), entonces  $p(\boldsymbol{\theta}|\mathbf{y})$  se puede obtener aproximadamente como el promedio de  $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$  sobre las  $\mathbf{z}$ 's imputadas. Sin embargo,  $p(\mathbf{z}|\mathbf{y})$  depende, a su vez, de  $p(\boldsymbol{\theta}|\mathbf{y})$ , por lo tanto si esta se conociera, podría usarse para calcular  $p(\mathbf{z}|\mathbf{y})$ . Esta dependencia mutua entre  $p(\boldsymbol{\theta}|\mathbf{y})$  y  $p(\mathbf{z}|\mathbf{y})$  conduce a un algoritmo iterativo para calcular  $p(\boldsymbol{\theta}|\mathbf{y})$ . Analíticamente, este algoritmo es esencialmente el método de sustitución sucesiva para resolver una ecuación de operador de punto fijo ([Tanner y Wong, 1987](#)).

El método consiste en iteraciones de los siguientes dos pasos:

- a) Dada la estimación actual de la distribución posterior de  $\boldsymbol{\theta}$  dado  $\mathbf{y}$ , generar una muestra de tamaño  $n > 0$  de datos latentes a partir de la distribución predictiva de  $\mathbf{z}$  dado  $\mathbf{y}$ ,
- b) Actualizar la posterior de  $\boldsymbol{\theta}$  dado  $\mathbf{y}$ , para ser la mezcla de los  $n$  datos aumentados.

Note que el foco de inferencia es:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int_{\mathbf{z}} p(\boldsymbol{\theta}|\mathbf{z}, \mathbf{y}) p(\mathbf{z}|\mathbf{y}) d\mathbf{z}$$

y esta marginalización se realiza a través de métodos de Cadenas de Markov Monte Carlo (MCMC) ([Sorensen y Gianola, 2007](#)), algunos de ellos, como lo son Muestreador de Gibbs y Algoritmo de Metropolis, presentados en secciones siguientes.

## 2.3. Inferencia Bayesiana

La inferencia Bayesiana es el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado mediante una distribución de probabilidad en los parámetros del modelo y en cantidades no observadas, como las predicciones para nuevas observaciones. La metodología Bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes. Bajo el enfoque Bayesiano, el vector de parámetros de interés,  $\boldsymbol{\theta}$ , sobre el cual se desea hacer algún tipo de inferencia es desconocido, de tal modo que el objetivo es usar datos,  $\mathbf{y}|\boldsymbol{\theta}$ , junto con un modelo paramétrico dado, para realizar análisis sobre los parámetros desconocidos.

El proceso de análisis de datos Bayesiano se puede visualizar dividiéndolo en los siguientes tres pasos:

1. Configuración de un modelo de probabilidad completo: una distribución de probabilidad conjunta para todas las cantidades observables y no observables en un problema.
2. Condicionamiento sobre los datos observados: calcular e interpretar la distribución posterior apropiada; la distribución de probabilidad condicional de las cantidades no observadas de interés final, dados los datos observados.
3. Evaluar el ajuste del modelo y las implicaciones de la distribución posterior resultante (Gelman *et al.*, 2014).

### El Teorema de Bayes

Con el fin de obtener información probabilística sobre  $\boldsymbol{\theta}$  dado  $\mathbf{y}$ , debemos comenzar con un modelo que proporcione una distribución de probabilidad conjunta para  $\boldsymbol{\theta}$  y  $\mathbf{y}$ . La función de masa o densidad de probabilidad conjunta puede escribirse como un producto de dos densidades que a menudo se conocen como la distribución a priori  $p(\boldsymbol{\theta})$  y la distribución de muestreo (o distribución de datos o función de verosimilitud)  $p(\mathbf{y}|\boldsymbol{\theta})$ ,

$$p(\boldsymbol{\theta}, \mathbf{y}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Simplemente condicionando el valor conocido de los datos  $\mathbf{y}$ , utilizando la propiedad básica de probabilidad condicional conocida como Teorema de Bayes, se obtiene la densidad posterior  $\boldsymbol{\theta}|\mathbf{y}$ , es decir:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})},$$

donde  $p(\mathbf{y})$ , se puede obtener de dos maneras:

## 2.4. Muestreador de Gibbs

---

$$p(\mathbf{y}) = \begin{cases} \sum_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) & \text{si } \boldsymbol{\theta} \text{ es discreto.} \\ \int_{\Theta} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} & \text{si } \boldsymbol{\theta} \text{ es continuo,} \end{cases}$$

donde  $\Theta$ , denota el espacio paramétrico de  $\boldsymbol{\theta}$  y  $p(\mathbf{y})$  es la constante de normalización de  $p(\boldsymbol{\theta}|\mathbf{y})$ , y es llamada la distribución marginal de los datos o la distribución predictiva inicial (Carlin y Louis, 2000). Entonces, como  $p(\mathbf{y})$  no depende de  $\boldsymbol{\theta}$ , se tiene la densidad posterior no normalizada:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

El primer término del lado derecho de la expresión anterior,  $p(\mathbf{y}|\boldsymbol{\theta})$ , es tomado como una función de  $\boldsymbol{\theta}$ , no de  $\mathbf{y}$ . Estas fórmulas simples encapsulan el núcleo de la inferencia Bayesiana, la tarea principal de cualquier aplicación específica es desarrollar el modelo  $p(\boldsymbol{\theta}, \mathbf{y})$  y realizar los cálculos para resumir  $p(\boldsymbol{\theta}|\mathbf{y})$  de manera apropiada (Gelman *et al.*, 2014).

Generalmente es difícil obtener  $p(\boldsymbol{\theta}|\mathbf{y})$ , de forma analítica, por lo cual es común utilizar algoritmos numéricos como Metropolis-Hastings (Hastings, 1970), descrito una sección posterior; muestreador de Gibbs (Geman y Geman, 1984), etc., este último se describe a continuación.

## 2.4. Muestreador de Gibbs

El muestreador de Gibbs (Geman y Geman, 1984) es una técnica de Cadenas de Markov Monte Carlo (MCMC), que permite muestrear de una distribución conjunta utilizando sus distribuciones condicionales.

Suponga que  $p(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^t$  es una densidad conjunta y se quiere obtener:

$$p(\theta_j) = \int \dots \int p(\theta_1, \theta_2, \dots, \theta_p) d\theta_1, d\theta_2, \dots, d\theta_{j-1}, d\theta_{j+1}, \dots, d\theta_p,$$

pero en muchos casos la integral anterior es muy compleja de obtener de forma analítica. Un algoritmo de cadena de Markov particular que se ha encontrado útil para resolver este problema es el muestreador de Gibbs, también llamado en ocasiones muestreo condicional alternativo.

Siguiendo a Sorensen y Gianola (2007), considere el vector de parámetros  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^t$ , cuya distribución posterior es proporcional a

## 2.4. Muestreador de Gibbs

$$p(\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_j, \theta_{j+1}, \dots, \theta_p | \mathbf{y})$$

Sea  $\boldsymbol{\theta}_{-j} = (\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p)^t$  un vector de dimensión  $(p-1)$ , que es igual a  $\boldsymbol{\theta}$  pero con su componente  $\theta_j$  eliminado. La distribución posterior condicional completa de  $\theta_j$  es:

$$\begin{aligned} p(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y}) &= \frac{p(\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_j, \theta_{j+1}, \dots, \theta_p | \mathbf{y})}{\int p(\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_j, \theta_{j+1}, \dots, \theta_p | \mathbf{y}) d\theta_j} \\ &\propto p(\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_j, \theta_{j+1}, \dots, \theta_p | \mathbf{y}) \end{aligned}$$

Note que también es posible obtener la distribución posterior condicional completa de  $\boldsymbol{\theta}_r$ , con  $r > 1$ , el número de elementos del vector de parámetros de interés ( $\boldsymbol{\theta}_r$ ) y además  $p > r$ . Sin embargo, en muchas aplicaciones  $r = 1$ , como en el caso descrito anteriormente, y los parámetros se actualizan de uno en uno al muestrear, cuando esto es posible, de  $p(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y})$ , o lo que comunmente se denota como  $p(\theta_j | \text{resto})$ .

Una explicación simple de cómo y por qué funciona el algoritmo de Gibbs puede encontrarse en [Casella y George \(1992\)](#). Los pasos necesarios para implementar el muestreador de Gibbs se detallan en el Algoritmo 2.

### Algoritmo 2: Muestreador de Gibbs.

```

input : Para generar una cadena de Markov, el muestreador de Gibbs comienza
          con un valor inicial  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$  en el soporte de  $p(\boldsymbol{\theta} | \mathbf{y})$ .

output: En el tiempo  $t$ , la realización  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$  se obtiene
          iterando en el ciclo siguiente.

1 initialization  $t = 0$ 
2 for  $t \leq B$  do /* con B número de muestras deseadas */
3   | Hacer  $t = t + 1$ .
4   | Se genera  $\theta_1^{(t)}$  de  $p(\theta_1 | \mathbf{y}, \theta_2^{(t-1)}, \dots, \theta_p^{(t-1)})$ .
5   | Se genera  $\theta_2^{(t)}$  de  $p(\theta_2 | \mathbf{y}, \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)})$ .
6   |  $\vdots$ 
7   | Se genera  $\theta_p^{(t)}$  de  $p(\theta_p | \mathbf{y}, \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{p-1}^{(t)})$ .
8 end

/* Las primeras  $m$  muestras se denominan, muestras de calentamiento
   (burn-in), así que se pueden desechar y solamente tomar en
   cuenta las últimas  $B - m$  muestras a las cuales se les aplica el
   adelgazamiento (thin), que consiste en guardar únicamente los
    $k$ -ésimos pasos de la cadena, lo que resulta en muestras con
   menos autocorrelación. */

```

## 2.5. Algoritmo de Metropolis

Este algoritmo es un método de Monte Carlo vía Cadenas de Markov (MCMC) el cual ha sido ampliamente utilizado en Física, restauración de imágenes y actualmente en Estadística (Carlin y Louis, 2000). El método en su versión original fue propuesto por Metropolis *et al.* (1953), y evoluciona al conocido algoritmo como Metropolis-Hastings Hastings (1970), una exposición detallada de este algoritmo se encuentra en Chib y Greenberg (1995).

Siguiendo a Gelman *et al.* (2014), considere el vector de parámetros  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^t$  cuya distribución  $p(\boldsymbol{\theta})$  es completamente conocida y de la cual se desea generar muestras aleatorias. Para construir la cadena  $\{\boldsymbol{\theta}^{(t)}\}$  se genera un punto candidato  $\boldsymbol{\theta}^*$  de una distribución propuesta o distribución generadora de candidatos en el tiempo  $t$ ,  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ . Para el algoritmo de Metropolis, la distribución generadora de candidatos debe ser simétrica, satisfaciendo la condición  $J_t(\boldsymbol{\theta}_a|\boldsymbol{\theta}_b) = J_t(\boldsymbol{\theta}_b|\boldsymbol{\theta}_a)$  para todo  $\boldsymbol{\theta}_a, \boldsymbol{\theta}_b$  y  $t$ .

Note que la distribución propuesta,  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ , solo depende del estado  $\boldsymbol{\theta}^{(t-1)}$ . Así el punto candidato  $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(t)}$  se acepta con la siguiente probabilidad de transición:

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^* & \text{con probabilidad } \alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}) = \min(r, 1), \\ \boldsymbol{\theta}^{(t-1)} & \text{d.o.m.} \end{cases}$$

donde

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})}.$$

Note que  $r$  es la razón con la cual se decide si se acepta o se rechaza el candidato. Si  $\boldsymbol{\theta}^*$  no se acepta, entonces la cadena no se mueve y  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ . Además la probabilidad de transición de  $\boldsymbol{\theta}^{(t-1)}$  a  $\boldsymbol{\theta}^{(t)}$ ,  $\alpha(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{(t-1)})$ , depende sólo de  $\boldsymbol{\theta}^{(t-1)}$  y no de  $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(t-2)}, \dots$  y cuya distribución de Transición se define como:

$$T_t(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}) = \alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t-1)}) J_t(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}).$$

Dado el valor actual  $\boldsymbol{\theta}^{(t-1)}$ , la distribución de transición  $T_t(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$  de la cadena de Markov es, por lo tanto, una mezcla de una masa puntual en  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ , y una versión ponderada de la distribución propuesta,  $J_t(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$ , que se ajusta a la tasa de aceptación.

## 2.6. Caso Caminata Aleatoria Metropolis (RWM)

Existen dos casos particulares de  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$  que comúnmente se utilizan y dan origen a las variantes más populares del algoritmo de Metropolis, *Caminata Aleatoria y Muestreo Independiente* (Smith y Roberts, 1993).

El Algoritmo 3 resume los pasos, descritos anteriormente, necesarios para implementar el Algoritmo de Metropolis.

**Algoritmo 3:** Algoritmo de Metropolis.

```

input : Para generar una cadena de Markov,  $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(B)})$ , el algoritmo
          de Metropolis comienza con un valor inicial  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ 
          en el soporte de  $p(\boldsymbol{\theta}|\mathbf{y})$ .
output: En el tiempo  $t$ , la realización  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$  se obtiene
          iterando en el ciclo siguiente.

1 initialization  $t = 0$ 
2 for  $t \leq B$  do /* con B número de muestras deseadas */
3   | Hacer  $t = t + 1$ .
4   | Generar una observación  $\boldsymbol{\theta}^*$  de la distribución generadora de candidatos
   |  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ .
5   | Generar una variable  $u \sim U(0, 1)$ .
6   | if  $u < r$  then
7   |   | hacer  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$ ,
8   |   | else
9   |   | hacer  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ .
10  | end
11 end

```

## 2.6. Caso Caminata Aleatoria Metropolis (RWM)

Sea  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = J_1(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(t-1)})$ , donde  $J_1(\cdot)$  es una densidad de probabilidades simétrica centrada en el origen. Entonces el punto candidato  $\boldsymbol{\theta}^*$  se genera con:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(t-1)} + \boldsymbol{\eta},$$

donde  $\boldsymbol{\eta}$  es una variable aleatoria de incremento de la distribución generadora de candidatos  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ , la cual se genera a partir de una distribución esférica simétrica, es decir, el caso Gaussiano, donde  $\boldsymbol{\eta} \sim N^{(p)}(\mathbf{0}, \boldsymbol{\Sigma})$ , y con base en lo anterior:

## 2.7. Caso Muestreo Independiente (IS)

---

$$\alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}) = \min(r, 1), \quad r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}|\mathbf{y})} \text{ (Craiu y Rosenthal, 2014).}$$

## 2.7. Caso Muestreo Independiente (IS)

Sea  $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = J_0(\boldsymbol{\theta}^*)$ , donde  $J_0(\cdot)$  es una densidad de probabilidad sobre  $\boldsymbol{\theta}^{(t-1)}$ , es decir,  $\boldsymbol{\theta}^*$  no depende del estado actual de la cadena,  $\boldsymbol{\theta}^{(t-1)}$ , por lo tanto:

$$\alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}) = \min(r, 1), \quad r = \frac{\omega(\boldsymbol{\theta}^*)}{\omega(\boldsymbol{\theta}^{(t-1)})},$$

con  $\omega(\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|\mathbf{y})}{J_0(\boldsymbol{\theta})}$  (Tierney, 1994).

En general, el RWM se utiliza en situaciones en las que tenemos poca idea sobre la forma de la distribución objetivo y, por lo tanto, debemos movernos a través del espacio de parámetros. La situación opuesta es una en la que tenemos una idea bastante buena de la distribución objetivo y podemos producir una aproximación creíble  $J_0(\cdot)$  que se puede usar como propuesta en el algoritmo IS.

En la práctica, se debe decidir qué muestra usar y, lo que es más importante, qué parámetros de simulación elegir. Por ejemplo, en el caso de RWM, la matriz de covarianza de la propuesta,  $\boldsymbol{\Sigma}$ , desempeña un papel crucial en la eficiencia del algoritmo de muestreo.

Las modificaciones de estos muestreadores incluyen además, el método de Rechazo Retrazado (DR), Metropolis de Múltiples Intentos (MTM) y los algoritmos de Saltos Reversibles (RJ), entre otros (Smith y Roberts, 1993).

## 2.8. Regresión con Errores aleatorios *Normales*

El análisis de regresión es una técnica estadística (posiblemente la más usada) para investigar y modelar la relación entre dos o más variables (Montgomery *et al.*, 2006). El primer método analítico para la estimación de la recta de regresión, el Método de Mínimos Cuadrados Ordinarios (MCO), se atribuye a M. Legendre (1805) con su obra “Nuevos métodos para la determinación de las órbitas de cometas”, mismo que se desarrolló de manera más profunda por Gauss (1809) en su obra “Teoría del movimiento de los cuerpos celestes sobre el sol en las secciones canónicas”. La idea del

## 2.8. Regresión con Errores aleatorios *Normales*

---

método es producir estimadores,  $\hat{\beta}_0$  y  $\hat{\beta}$  que minimicen la suma de cuadrados de las distancias entre los valores observados  $y_i$  y los estimados  $\hat{y}_i$ . Además de MCO, pueden utilizarse otros métodos para resolver el problema de regresión lineal, como lo son, estimación por Máxima Verosimilitud (EMV) y Estimación Bayesiana, dependiendo de la clase de modelo lineal presentado y el tipo de inferencias que se deseen acerca de los parámetros.

A continuación se presenta formalmente el modelo clásico de regresión lineal (MCRL). Sean  $Y_1, \dots, Y_n$  variables aleatorias observables no correlacionadas entre si y,  $x_{i1}, \dots, x_{ip}$ ,  $i = 1, \dots, n$ , un conjunto de observaciones (a  $Y_i$  se le llama comúnmente variable respuesta y a  $x_{ij}$ ,  $j = 1, \dots, p$ , variables predictoras o covariables), el MCRL se define por la relación:

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\ e_i \sim iidN(0, \sigma^2),$$

donde  $\beta_0$  y  $\boldsymbol{\beta}$  son parámetros de regresión desconocidos, el primero representa la ordenada al origen de la recta de regresión y no siempre tiene interpretación práctica, pero es necesario para representar la tendencia que muestran los datos en el espacio de valores observados para las covariables, el segundo indica la forma en que están relacionadas  $\mathbf{X}$  y  $\mathbf{Y}$ , en el sentido en que mide cuánto y en qué dirección se modifican los valores de  $\mathbf{Y}$  cuando cambia  $\mathbf{X}$ . Y  $e_i$  son variables aleatorias no observables e independientes con distribución *Normal* cuya media se supone 0 y varianza  $\sigma^2$ .

De lo anterior, por propiedades de la distribución Normal, es posible demostrar que

$$Y_i \sim N(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \sigma^2), \text{ independientes.}$$

De este modo, con la declaración de los supuestos fundamentales para el MCRL, se dice que el modelo está identificado (Mood, 1974).

Existen muchas propiedades deseables que se distinguen en el MCRL, sin embargo, debido a la necesidad de adoptar los supuestos antes mencionados, existen también situaciones en que este modelo no resulta adecuado. De esta manera se tiene que el MCRL resulta ser muy sensible a observaciones alejadas del resto de los datos, es decir, una sola observación apartada del resto puede modificar sustancialmente la estimación realizada. A este tipo de observaciones se les conoce como valores atípicos u observaciones aberrantes que es un caso que no sigue el mismo modelo que el resto de los datos. Sin embargo, debe señalarse que esta deficiencia en las estimaciones ante la presencia de este tipo de observaciones no proviene del modelo lineal, sino del método de estimación elegido para ajustarlo: el método de MCO que propone como función de pérdida a la suma del cuadrado de los residuos. Los métodos de estimación que son menos sensibles a la influencia de observaciones apartadas del resto se denominan “métodos robustos de estimación”. Si en vez de tomar esa función de pérdida, se eligiera

## 2.9. Regresión con Errores aleatorios *t-Student*

---

otra, se podría subsanar este déficit que tienen los estimadores de MCO, de volverse ilimitadamente sensibles a una observación atípica. Sin embargo, esta solución funciona siempre y cuando los datos de la muestra siguieran el modelo de regresión con errores *Normales* (Szretter, 2017). Pero, ¿qué sucede cuando existe una cantidad considerable de datos que pueden considerarse como valores atípicos u observaciones aberrantes en el MCRL? Se rompe el supuesto de normalidad en los errores y se puede incurrir así en un problema de especificación del modelo, lo que da paso a modelos de regresión robusta que asumen diferentes distribuciones a la *Normal* en los errores, es decir, que ofrecen un mejor ajuste y permiten la obtención de estimaciones más adecuadas y confiables.

## 2.9. Regresión con Errores aleatorios *t-Student*

Como ya se mencionó, se sabe que la inferencia estadística basada en la distribución *Normal* (univariada o multivariada) es sensible a la presencia de valores atípicos. A pesar de este hecho, y del considerable interés en procedimientos robustos, en la literatura estadística matemática, la mayoría de los análisis estadísticos aplicados continúan basándose en el modelo *Normal*. Muchos problemas quedan fuera del rango de los modelos convencionales, y los modelos deben elegirse para que se ajusten a la ciencia y los datos subyacentes, no simplemente por su conveniencia analítica o computacional. La distribución de probabilidades *t*, propuesta por W.S. Gosset en 1908, proporciona una extensión útil de lo *Normal* para el modelado estadístico de conjuntos de datos que implican errores con colas más largas que las *Normales* (Gelman et al., 2014). En el marco Bayesiano, es posible hacer inferencias a través de la obtención de muestras de la distribución posterior; en esta sección se presentan los desarrollos teórico-matemáticos necesarios para el uso del muestreador de Gibbs en la formulación de un modelo robusto, teniendo a los errores de regresión distribuidos como una *t-Student*.

### 2.9.1. Distribución *t-Student*

Sea  $Y$  una variable aleatoria con función de densidad:

$$p(y|\mu, \sigma^2, k) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2) \sqrt{k\pi\sigma^2}} \left[ 1 + \frac{(y-\mu)^2}{k\sigma^2} \right]^{-\frac{(k+1)}{2}}, \quad y \in \mathbb{R} \quad (2.1)$$

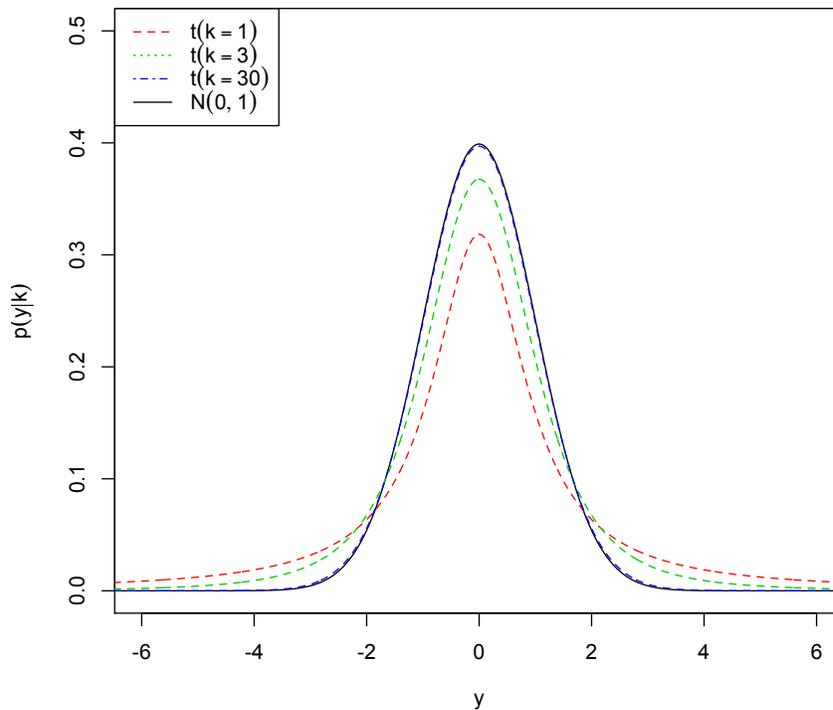
donde  $\mu \in \mathbb{R}$  parámetro de localización,  $\sigma^2 > 0$  parámetro de escala,  $k > 0$  grados de libertad y  $\Gamma(\cdot)$  es la función gamma,  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$  con  $z > 0$ . La función de densidad en (2.1) corresponde a la de una variable aleatoria  $t(\mu, \sigma^2, k)$  (Sorensen y Gianola, 2007). A

## 2.9. Regresión con Errores aleatorios *t-Student*

continuación se enumeran algunas de sus propiedades:

1. Esta distribución es simétrica respecto al parámetro de localidad  $\mu$ .
2. Cuando  $k \leq 2$  entonces  $\text{var}(y) = \infty$ .
3. Si  $\mu = 0$  y  $\sigma^2 = 1$  se tiene la distribución  $t(0, 1, k)$  o simplemente  $t(k)$ .
4. Cuando  $k \rightarrow \infty$  la distribución converge a una  $N(\mu, \sigma^2)$ .
5. Cuando  $k = 1$  se tiene una distribución Cauchy,  $C(\mu, \sigma^2)$ , cuya media y momentos de orden superior no existen.

En la Figura 2.1 se puede observar la densidad de la *t-Student* con  $\mu = 0$  y  $\sigma^2 = 1$ , para distintos valores de  $k$ , grados de libertad. Algunas de las propiedades enunciadas se pueden identificar con facilidad en esta figura.



**Figura 2.1:** Densidad de la  $t(k)$  con distintos grados de libertad,  $k$ . Observe que si  $k \rightarrow \infty$  se obtiene la densidad  $N(0, 1)$  y si  $k = 1$  se obtiene la densidad  $C(0, 1)$ .

## 2.9. Regresión con Errores aleatorios *t-Student*

### 2.9.2. Regresión Bayesiana con Errores *t-Student*

Si se considera el modelo de regresión lineal múltiple

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\ e_i \sim iidt(0, \sigma^2, k),$$

donde  $t(0, \sigma^2, k)$  denota la función de densidad de probabilidades *t-Student* dada en (2.1).

Una representación equivalente para este modelo de regresión es:

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + \frac{\varepsilon_i}{\sqrt{\omega_i}}, \quad i = 1, 2, \dots, n; \\ \varepsilon_i \sim iidN(0, \sigma^2), \\ \omega_i \sim Ga \left( shape = \frac{k}{2}, rate = \frac{k}{2} \right).$$

El planteamiento anterior proviene del punto de vista clásico para la generación de una variable aleatoria con distribución *t-Student* (Sorensen y Gianola, 2007). La forma de probar que la razón de las variables aleatorias  $\varepsilon_i$  y  $\sqrt{\omega_i}$  dan origen a otra variable aleatoria con distribución *t-Student* consiste en encontrar la distribución conjunta de  $\varepsilon_i$  y  $\omega_i$ ,  $p(\varepsilon_i, \omega_i)$ , definir la transformación  $\frac{\varepsilon_i}{\sqrt{\omega_i}}$  y obtener su distribución a partir de la primera, utilizando Jacobianos (Casella y Berger, 2002).

Note que el vector de parámetros incluye los coeficientes de regresión  $\beta_0$ ,  $\boldsymbol{\beta}$  y el parámetro  $\sigma^2$ , es decir,  $\boldsymbol{\theta} = [\beta_0, \boldsymbol{\beta}, \sigma^2]^t$  con  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t$ , mientras que alternativamente  $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_n]^t$  son considerados pesos, variables no observables (latentes). Esto se conoce como “Aumentación de datos” (Tanner y Wong, 1987). Bajo esta consideración puede demostrarse que  $y$  se distribuye como una *t-Student* de la siguiente manera:

$$y_i | \omega_i, \boldsymbol{\theta} \sim N \left( \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \frac{\sigma^2}{\omega_i} \right) \\ \omega_i | k \sim Ga \left( shape = \frac{k}{2}, rate = \frac{k}{2} \right),$$

la distribución conjunta de  $y_i, \omega_i$  queda definida como sigue:

$$p(y_i, \omega_i | \boldsymbol{\theta}, k) = p(y_i | \omega_i, \boldsymbol{\theta}) p(\omega_i | k),$$

## 2.9. Regresión con Errores aleatorios *t-Student*

---

entonces, la distribución de  $Y_i|\boldsymbol{\theta}, k$  se obtiene por integración:

$$p(y_i|\boldsymbol{\theta}, k) = \int p(y_i, \omega_i|\boldsymbol{\theta}, k) d\omega_i,$$

resolviendo la integral anterior se puede mostrar que:

$$Y_i \sim t(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \sigma^2, k), \text{ independientes.}$$

La Verosimilitud Aumentada se obtiene como sigue:

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, k) &= \prod_{i=1}^n p(y_i, \omega_i|\boldsymbol{\theta}, k) \\ &= \prod_{i=1}^n \left[ \frac{\sqrt{\omega_i}}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{\omega_i}{2\sigma^2} [y_i - (\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta})]^2 \right\} \times p \left( \omega_i \middle| \frac{k}{2}, \frac{k}{2} \right) \right]. \end{aligned}$$

Suponga las distribuciones a priori:

$$\begin{aligned} \beta_0 | \sigma_\beta^2 &\sim N(0, \sigma_\beta^2), \\ \boldsymbol{\beta} | \sigma_\beta^2 &\sim NM(\mathbf{0}, \sigma_\beta^2 \mathbf{I}), \\ p(\sigma^2) &\propto \frac{1}{\sigma^2}. \end{aligned}$$

Cabe señalar que para la distribución a priori de  $\beta_0$  y  $\boldsymbol{\beta}$  se fijó  $\sigma_\beta^2 = 1 \times 10^6$  dando lugar a una distribución a priori no informativa. Además, el parámetro  $k$  se considera como fijo.

Con el fin de obtener estimaciones de  $\boldsymbol{\theta}$  aplicamos el Teorema de Bayes utilizando la función de verosimilitud aumentada, la distribución a priori de  $\boldsymbol{\theta}$  y  $\boldsymbol{\Omega}$  el conjunto de hiperparámetros y el muestreador de Gibbs.

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\omega}|\text{datos}) &\propto p(\mathbf{y}, \boldsymbol{\omega}|\boldsymbol{\theta}, k) p(\boldsymbol{\theta}|\boldsymbol{\Omega}) \\ &= \prod_{i=1}^n [p(y_i, \omega_i|\boldsymbol{\theta}, k)] p(\beta_0) p(\boldsymbol{\beta}) p(\sigma^2). \end{aligned}$$

## 2.9. Regresión con Errores aleatorios *t-Student*

Entonces utilizando la función de verosimilitud de los datos aumentados y las densidades a priori antes definidas, considerando además a  $k$  como fijo, la distribución a posteriori  $p(\boldsymbol{\theta}, \boldsymbol{\omega} | \text{datos})$  está dada por:

$$p(\boldsymbol{\theta}, \boldsymbol{\omega} | \text{datos}) \propto \prod_{i=1}^n \left( \frac{k}{2} \right)^{\frac{k}{2}} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left( \frac{\sigma^2}{\omega_i} \right)^{-\frac{1}{2}} \omega_i^{\frac{k}{2}-1} \exp \left\{ -\frac{\omega_i}{2} \left[ \frac{(y_i - \beta_0 - \mathbf{x}_i^t \boldsymbol{\beta})^2 + k\sigma^2}{\sigma^2} \right] \right\} \\ \times N(\beta_0 | 0, \sigma_\beta^2) NM(\boldsymbol{\beta} | \mathbf{0}, \sigma_\beta^2 \mathbf{I}) \frac{1}{\sigma^2}. \quad (2.2)$$

Considere además que  $\hat{e}_i = y_i - (\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta})$  son los errores de regresión.

Entonces las distribuciones condicionales completas para implementar el algoritmo de muestreo de Gibbs se pueden obtener de (2.2) y se presentan a continuación (Sorensen y Gianola, 2007):

1.  $p(\omega_i | \text{resto}), i = 1, \dots, n$

$$\omega_i | \text{resto} \sim Ga \left( \frac{k+1}{2}, rate = \frac{k + \hat{e}_i^2 / \sigma^2}{2} \right).$$

2.  $p(\beta_0 | \text{resto})$

$$\beta_0 | \text{resto} \sim N \left( \frac{\sum_{i=1}^n \omega_i y_i^*}{\sum_{i=1}^n \omega_i + \lambda}, \frac{\sigma^2}{\sum_{i=1}^n \omega_i + \lambda} \right), \quad y_i^* = \hat{e}_i + \beta_0^{old}, \quad \lambda = \frac{\sigma^2}{\sigma_\beta^2}.$$

3.  $p(\beta_j | \text{resto}), j = 1, \dots, p$

$$\beta_j | \text{resto} \sim N \left( \frac{\sum_{i=1}^n \omega_i y_i^{**} x_{ij}}{\sum_{i=1}^n \omega_i x_{ij}^2 + \lambda}, \frac{\sigma^2}{\sum_{i=1}^n \omega_i x_{ij}^2 + \lambda} \right), \quad y_i^{**} = \hat{e}_i + \beta_j^{old} x_{ij}, \quad \lambda = \frac{\sigma^2}{\sigma_\beta^2}.$$

4.  $p(\sigma^2 | \text{resto})$

$$\sigma^2 | \text{resto} \sim GI \left( shape = \frac{n}{2}, rate = \frac{SCE}{2} \right), \quad SCE = \sum_{i=1}^n \omega_i \hat{e}_i^2.$$

## 2.9. Regresión con Errores aleatorios *t-Student*

Note que  $GI(\cdot)$  denota una variable aleatoria *Gamma Invertida* (Sorensen y Gianola, 2007).

Para fijar el parámetro  $k$  se sugiere utilizar una malla de puntos, por ejemplo  $k \in (2, 30)$ , y elegir el mejor modelo comparando el valor del Criterio de Información de Akaike (AIC).

Los pasos necesarios para el funcionamiento del muestreador de Gibbs para ajustar un modelo de regresión con errores aleatorios *t-Student* se describen en el Algoritmo 4.

<b>Algoritmo 4:</b> Regresión Bayesiana con errores aleatorios $t$ .	
<b>input</b>	: El algoritmo requiere que se le proporcione $\mathbf{y}$ , $\mathbf{X}$ y fijar $k$ .
<b>output:</b>	$p(\boldsymbol{\theta}, \boldsymbol{\omega}   \text{resto})$ .
1	<b>for</b> $h$ in $1 : B$ <b>do</b> /* con $B$ número de iteraciones */
2	Inicializar $\boldsymbol{\theta}^{old} = (\beta_0^{old} = 0, \beta_j^{old} = 0, \sigma^{2old} = s_y^2)$ y fijar $\sigma_\beta^2 = 1 \times 10^6$ .
3	Obtener $\hat{e}_i = y_i - (\beta_0^{old} + \mathbf{x}_i^t \boldsymbol{\beta}^{old})$ , $i = 1, \dots, n$ .
4	Muestrar de cada distribución posterior $\boldsymbol{\theta}   \text{resto}$ , obtener $\boldsymbol{\theta}^{new}$ , y actualizar $\hat{\mathbf{e}}$ tras actualizar cada coeficiente de regresión $\beta_0$ y $\beta_j$ , $j = 1, \dots, p$ .
5	<b>begin</b>
6	$\omega_i   \beta_0^{old}, \beta_j^{old}, \sigma^{2old} \rightarrow \omega_i^{new}$
7	$\beta_0   \omega_i^{new}, \beta_j^{old}, \sigma^{2old} \rightarrow \beta_0^{new}$
8	$\delta = \beta_0^{old} - \beta_0^{new}$
9	$\hat{e}_i = \hat{e}_i + \delta$
10	<b>for</b> $j = 1, \dots, p$ <b>do</b>
11	$\beta_j   \omega_i^{new}, \beta_0^{new}, \sigma^{2old} \rightarrow \beta_j^{new}$
12	$\delta = \beta_j^{old} - \beta_j^{new}$
13	$\hat{e}_i = \hat{e}_i + \delta x_{ij}$
14	<b>end</b>
15	$\sigma^2   \omega_i^{new}, \beta_0^{new}, \beta_j^{new} \rightarrow \sigma^{2new}$
16	<b>end</b>
17	<b>end</b>
	/* Las primeras $m$ muestras son el <i>calentamiento</i> (burn-in) así que se pueden desechar y solamente tomar en cuenta las últimas $B - m$ muestras a las cuales se les aplica el <i>adelgazamiento</i> (thin). */

El Algoritmo 4 se implementó en el paquete estadístico *R* (R Core Team, 2018) y se presenta en el Anexo A.

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

### 2.10.1. Distribución *Normal-Asimétrica*

En el contexto de los modelos lineales, existen muchas aplicaciones reales donde los supuestos de normalidad y homocedasticidad no se satisfacen. Ocurre que en muchos casos, la distribución de frecuencias asociada a la variable respuesta es marcadamente asimétrica. [Azzalini \(1985\)](#) introduce el concepto de la *Distribución Normal-Asimétrica*, que permite modelar la asimetría presente en un conjunto de datos. Su función de densidad de probabilidades está dada por:

$$p(u|\lambda) = 2\phi(u) \Phi(\lambda u) I_{(-\infty, \infty)}(u), \quad (2.3)$$

donde  $\phi(\cdot)$  y  $\Phi(\cdot)$  denotan la función de densidad y distribución de una variable Normal estándar  $N(0, 1)$ , respectivamente, y  $\lambda \in \mathbb{R}$  regula la asimetría (forma).

Si  $U$  tiene la función de densidad dada en (2.3), entonces se escribe  $U \sim SN(\lambda)$ . Algunas propiedades de la distribución  $SN$ , que la hacen matemáticamente tratable son:

- 1) Si  $\lambda = 0$  en (2.3) entonces  $U \sim N(0, 1)$ .
- 2) Si  $U \sim SN(\lambda)$  entonces  $-U \sim SN(-\lambda)$ .
- 3) Si  $\lambda \rightarrow \infty$  entonces (2.3) converge a la distribución Media-Normal,  $2\phi(u)I_{(0, \infty)}(u)$ .
- 4) Si  $U \sim SN(\lambda)$  entonces  $U^2 \sim \chi_1^2$ .

Usando los resultados de [Azzalini \(1985\)](#), la función generadora de momentos de (2.3) está dada por:

$$M_U(t) = 2\exp(t^2/2)\Phi\left(\frac{\lambda t}{\sqrt{1 + \lambda^2}}\right),$$

de donde es posible obtener la media y la varianza, dados por:

$$E(U) = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1 + \lambda^2}}, \quad Var(U) = 1 - \frac{2}{\pi} \frac{\lambda^2}{1 + \lambda^2}.$$

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

---

y el coeficiente de asimetría de  $U$  está dado por:

$$\gamma_1 = \frac{[E(U)]^3}{\{1 - [E(U)]^2\}^{3/2}}.$$

La función de distribución acumulada de (2.3) está dada por:

$$P(U \leq u) = \int_{-\infty}^u \frac{1}{\pi} \exp(-x^2/2) \int_{-\infty}^{\lambda x} \exp(-t^2/2) dt dx.$$

Cosidere una transformación lineal de la variable aleatoria  $U$ ,  $Y = \xi + \omega U$ , donde  $\xi$  es un parámetro de localidad y  $\omega$  es un parámetro de escala, se dice entonces que la variable aleatoria  $Y \sim SN_D(\xi, \omega, \lambda)$ , donde el subíndice  $D$  representa la parametrización directa de la  $SN$ . Así la función de densidad de  $Y$  está dada por:

$$p(y|\xi, \omega, \lambda) = 2 \frac{1}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left[\lambda \left(\frac{y - \xi}{\omega}\right)\right], \quad (2.4)$$

donde  $y \in \mathbb{R}$ ,  $\xi \in \mathbb{R}$ ,  $\omega > 0$ . Se puede demostrar que el coeficiente de asimetría de  $Y$  corresponde al coeficiente de asimetría de  $U$ .

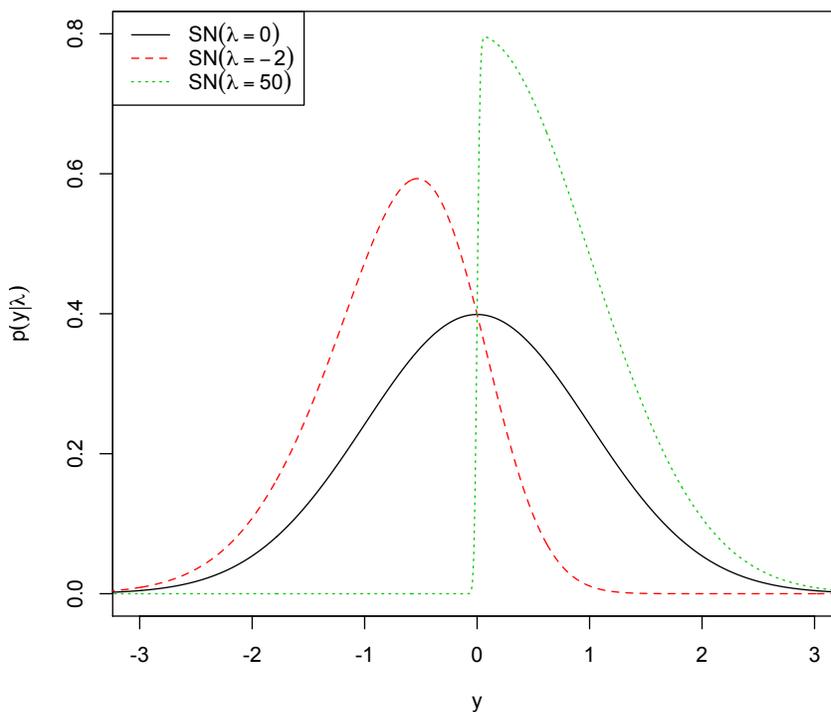
En la Figura 2.2 se puede observar la densidad de la *Normal Asimétrica* para  $\xi = 0$ ,  $\omega = 1$  y distintos valores de  $\lambda$ . Algunas de las propiedades enunciadas se pueden identificar con facilidad en la Figura 2.2.

Esta distribución tiene algunas propiedades de la distribución Normal univariada, pero presenta problemas desde el punto de vista de la estimación de sus parámetros, en particular, el estimador de máxima verosimilitud para el parámetro de forma tiende a infinito con probabilidad positiva (Azzalini, 1985; Pérez-Rodríguez *et al.*, 2017).

### 2.10.2. Regresión Bayesiana con Errores *Normal-Asimétricos* (*BSN*)

Si se considera el modelo de regresión lineal propuesto por Azzalini y Capitanio (1999) y Rusell-Noriega y González-Farías (2002)

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\ e_i \sim iidSN_D(0, \omega, \lambda),$$



**Figura 2.2:** Densidad de la  $SN(\lambda)$  con distintos parámetros de forma  $\lambda$ . Observe que si  $\lambda = 0$  se obtiene la densidad  $N(0,1)$  y si  $\lambda \rightarrow \infty$  se obtiene la distribución Media Normal.

y usando las propiedades de la distribución Normal Asimétrica, se sigue que

$$Y_i \sim SN_D(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \omega, \lambda).$$

Existen problemas para la estimación de los parámetros  $\xi$ ,  $\omega$ ,  $\lambda$ . Desde el punto de vista clásico, los estimadores de momentos no dan buenos resultados, y en el caso de Máxima Verosimilitud, existen diferentes problemas al momento de maximizar la función de verosimilitud (Pérez-Rodríguez *et al.*, 2018).

Azzalini y Capitanio (1999) proponen utilizar una reparametrización de la función de densidad (2.3), la cual se denomina *parametrización centrada*.

Si  $U \sim SN_D(\lambda)$  y  $\lambda = \frac{\rho}{\sqrt{1-\rho^2}}$ , con  $\rho \in (-1, 1)$ , entonces:

$$E(U) = \sqrt{\frac{2}{\pi}}\rho, \quad Var(U) = 1 - \frac{2}{\pi}\rho^2, \quad S(U) = \sqrt{Var(U)} = \sqrt{1 - \frac{2}{\pi}\rho^2}$$

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

---

se puede observar además que  $\rho = \frac{\lambda}{\sqrt{1-\lambda^2}}$ .

La variable aleatoria:

$$Y = \mu + \sigma_e \left( \frac{U - E(U)}{S(U)} \right),$$

sigue una distribución *SN* con parámetros  $\mu \in \mathbb{R}$  y  $\sigma_e > 0$ , los cuales denotan la media y la desviación estándar de  $Y$  respectivamente, es decir, en este caso  $E(Y) = \mu$  y  $Var(Y) = \sigma_e^2$ , entonces se dice que  $Y \sim SN_C(\mu, \sigma_e, \gamma_1)$ , donde  $\gamma_1 \in (-0.99527, 0.99527)$  es el coeficiente de asimetría dado por:

$$\gamma_1 = \sqrt{\frac{2}{\pi}} \rho^3 \left( \frac{4}{\pi} - 1 \right) \left( 1 - \frac{2\rho^2}{\pi} \right)^{-3/2}.$$

Considere la distribución  $p(u, z|\rho) = p(u|z, \rho) p(z)$  con  $p(u|z, \rho)$  la función de densidad de una variable aleatoria *Normal* con media  $\rho z$  y varianza  $1 - \rho^2$  y  $p(z)$  la función de densidad de una variable aleatoria *Normal Truncada*, con parámetro de localidad 0, parámetro de escala 1, cota inferior de truncamiento 0 y cota superior de truncamiento  $\infty$ . La distribución conjunta de  $U$  y  $Z$  es:

$$\begin{aligned} p(u, z|\rho) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (u - \rho z)^2 \right\} \\ &\times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z^2 \right\} I_z(0, \infty). \end{aligned} \tag{2.5}$$

Observe que la distribución de  $U$  puede ser obtenida por integración  $\int p(u, z|\rho) dz = p(u|\rho)$ . Para más detalles ver [Kim \(2005\)](#) y [Pérez-Rodríguez et al. \(2018\)](#) en la sección *Truncamiento oculto*.

Siguiendo el esquema de la parametrización centrada, considere la transformación

$$\begin{aligned} Y &= \mu + \sigma_e \left( \frac{U - E_U}{S_U} \right) \\ T &= Z \end{aligned}$$

con  $y \in \mathbb{R}$ ,  $z \geq 0$ ,  $E_U = E(U)$  y  $S_U = S(U)$ .

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

La distribución conjunta de las variables aleatorias  $Y$  y  $Z$  puede ser obtenida usando Jacobianos ([Casella y Berger, 2002](#)), y está dada por:

$$\begin{aligned}
 p(y, z | \mu, \sigma_e^2, \rho) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \frac{S_U}{\sigma_e} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ S_U \left( \frac{y-\mu}{\sigma_e} \right) + E_U - \rho z \right]^2 \right\} \\
 &\times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z^2 \right\} I_{(0,\infty)}(z).
 \end{aligned} \tag{2.6}$$

Note que la función de densidad de  $Y \sim SN_C(\mu, \sigma_e^2, \gamma_1)$  puede ser obtenida integrando  $p(y, z | \mu, \sigma_e^2, \rho)$  con respecto de  $z$ ; esto es

$$p(y | \mu, \sigma_e^2, \gamma_1) = \int_0^\infty p(y, z | \mu, \sigma_e^2, \rho) dz,$$

con  $\gamma_1 = \sqrt{\frac{2}{\pi}} \rho^3 \left( \frac{4}{\pi} - 1 \right) \left( 1 - \frac{2\rho^2}{\pi} \right)^{-3/2}$ . Esta representación es muy conveniente porque nos permite escribir una función de probabilidad aumentada ([Kim, 2005](#); [Liseo y Parisi, 2013](#)), como si hubiésemos observado los valores latentes  $z$ . La función de densidad de la variable aleatoria  $SN$  bajo la parametrización centrada es una función complicada que puede consultarse en [Azevedo et al. \(2011\)](#).

Entonces, considerando el modelo de regresión lineal presentado anteriormente pero ahora con los errores aleatorios  $SN_C$  se tiene

$$\begin{aligned}
 y_i &= \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\
 e_i &\sim iid SN_C(0, \sigma_e^2, \gamma_1),
 \end{aligned}$$

por lo tanto, por propiedades de la distribución *Normal Asimétrica* se tiene que

$$Y_i \sim SN_C(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \sigma_e^2, \gamma_1) \text{ independientes.}$$

Con esta parametrización centrada de la distribución  $SN$  en los errores, ahora el trabajo se enfoca en obtener estimaciones de los parámetros  $\boldsymbol{\theta} = [\beta_0, \boldsymbol{\beta}, \sigma_e^2, \gamma_1]^t$ ,  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t$ , con  $p(\boldsymbol{\theta} | \boldsymbol{\Omega})$  la distribución a priori de  $\boldsymbol{\theta}$  y  $\boldsymbol{\Omega}$  el conjunto de hiperparámetros, entonces se aplica el Teorema de Bayes y se obtiene la distribución posterior  $p(\boldsymbol{\theta} | \text{datos})$  como sigue:

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

---

$$\begin{aligned}
 p(\boldsymbol{\theta}|\text{datos}) &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\Omega) \\
 &= \prod_{i=1}^n [p(y_i|\boldsymbol{\theta})]p(\beta_0)p(\boldsymbol{\beta})p(\sigma_e^2)p(\gamma_1).
 \end{aligned}$$

Se tiene que ni la distribución posterior ni la distribución condicional de los parámetros de interés tiene forma cerrada, por lo tanto, para realizar la estimación [Kim \(2005\)](#) y [Pérez-Rodríguez et al. \(2018\)](#) proponen aumentación de datos lo cual permite utilizar el muestreador de Gibbs y el Algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de la distribución posterior.

Considere la distribución conjunta de  $Y$  y  $Z$  dada en (2.6). En el contexto de regresión, establecemos  $\mu_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}$ , entonces la función de verosimilitud aumentada es:

$$\begin{aligned}
 p(\mathbf{y}, \mathbf{z}|\beta_0, \boldsymbol{\beta}, \sigma_e^2, \rho) &= \prod_{i=1}^n p(y_i, z_i|\mu_i, \sigma_e^2, \rho) \\
 &\propto \prod_{i=1}^n \frac{1}{\sqrt{(1-\rho^2)}} \frac{S_U}{\sigma_e} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ S_U \left( \frac{y_i - \mu_i}{\sigma_e} \right) + E_U - \rho z_i \right]^2 - \frac{1}{2} z_i^2 \right\} I_{(0,\infty)}(z_i).
 \end{aligned}$$

Para especificar completamente el modelo Bayesiano, se deben definir distribuciones a priori para los parámetros desconocidos:

$$\begin{aligned}
 \beta_0|\sigma_\beta^2 &\sim N(0, \sigma_\beta^2), \\
 \boldsymbol{\beta}|\sigma_\beta^2 &\sim NM(\mathbf{0}, \sigma_\beta^2 \mathbf{I}), \\
 p(\sigma_e^2) &\propto \frac{1}{\sigma_e^2}.
 \end{aligned}$$

Cabe señalar que para la distribución a priori de  $\beta_0$  y  $\boldsymbol{\beta}$  se fijó  $\sigma_\beta^2 = 1 \times 10^6$  dando lugar a una distribución a priori no informativa.

Para el parámetro  $\rho$ , basados en la transformación  $\rho = 1 - 2B$ , donde  $B \sim \text{Beta}(a_0, b_0)$ , la densidad a priori queda definida como:

$$p(\rho|a_0, b_0) \propto \left( \frac{1-\rho}{2} \right)^{a_0-1} \left( 1 - \frac{1-\rho}{2} \right)^{b_0-1} I_{(-1,1)}(\rho).$$

Entonces utilizando la función de verosimilitud de los datos aumentados y las densidades

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

a priori antes definidas, la distribución a posteriori  $p(\boldsymbol{\theta}, \mathbf{z}|\text{datos})$  está dada por:

$$\begin{aligned}
 & p(\boldsymbol{\theta}, \mathbf{z}|\text{datos}) \\
 \propto & \prod_{i=1}^n \frac{1}{\sqrt{(1-\rho^2)}} \frac{S_U}{\sigma_e} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ S_U \left( \frac{y_i - \mu_i}{\sigma_e} \right) + E_U - \rho z_i \right]^2 - \frac{1}{2} z_i^2 \right\} I_{(0,\infty)}(z_i) \\
 \times & N(\beta_0|0, \sigma_\beta^2) NM(\boldsymbol{\beta}|\mathbf{0}, \sigma_\beta^2 \mathbf{I}) \frac{1}{\sigma_e^2} \times \left( \frac{1-\rho}{2} \right)^{a_0-1} \left( 1 - \frac{1-\rho}{2} \right)^{b_0-1} I_{(-1,1)}(\rho).
 \end{aligned} \tag{2.7}$$

Sea

$$\begin{aligned}
 \hat{e}_i &= y_i - \mu_i, \quad \mu_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \\
 \tau &= \frac{\sigma_e E_U}{S_U} - \frac{\sigma_e \rho z_i}{S_U},
 \end{aligned}$$

entonces, las distribuciones condicionales completas que se utilizan para la implementación del Muestreador de Gibbs y Algoritmo RWM del modelo *BSN* se obtienen de (2.7) y se presentan a continuación:

1.  $p(\beta_0|\text{resto})$

$$\beta_0|\text{resto} \sim N \left( \frac{\sum_{i=1}^n y_i^*}{n + \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \sigma_\beta^2}}, \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \left( n + \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \sigma_\beta^2} \right)} \right), \quad y_i^* = \hat{e}_i + \beta_0^{\text{old}} + \tau.$$

2.  $p(\beta_j|\text{resto}), j = 1, \dots, p$

$$\beta_j|\text{resto} \sim N \left( \frac{\sum_{i=1}^n x_{ij} y_i^*}{\sum_{i=1}^n x_{ij}^2 + \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \sigma_\beta^2}}, \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \left[ \sum_{i=1}^n x_{ij}^2 + \frac{(1-\rho^2)\sigma_e^2}{S_U^2 \sigma_\beta^2} \right]} \right), \quad y_i^* = \hat{e}_i + \beta_j^{\text{old}} x_{ij} + \tau.$$

3.  $p(z_i|\text{resto})$

$$p(z_i|resto) \propto \exp \left\{ -\frac{1}{2(1-\rho^2)} (z_i - \rho y_i^*)^2 \right\} I_{(0,\infty)}(z_i), \quad i = 1, \dots, n;$$

donde  $y_i^* = \frac{\hat{e}_i}{\sigma_e} S_U + E_U$ .

Se reconoce el núcleo de una distribución Normal Truncada con parámetro de localidad  $\rho y_i^*$ , parámetro de escala  $1 - \rho^2$ , cota inferior de truncamiento 0 y cota superior de truncamiento  $\infty$ , es decir,  $z_i \sim NT(\rho y_i^*, 1 - \rho^2, 0, \infty)$ .

#### 4. $p(\sigma_e^2|resto)$

En el caso de  $\sigma_e^2$  no es posible identificar el núcleo de alguna distribución univariada conocida, pues su distribución condicional no tiene forma cerrada:

$$p(\sigma_e^2|resto) \propto (\sigma_e^2)^{-\left(\frac{n}{2}+1\right)} \exp \left\{ -\frac{S_U^2}{2(1-\rho^2)\sigma_e^2} \sum_{i=1}^n \left( \hat{e}_i + \frac{\sigma_e E_U}{S_U} - \frac{\sigma_e \rho z_i}{S_U} \right)^2 \right\}.$$

Kim (2005) propone utilizar el algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de  $\sigma_e^2$ . Dado que  $\sigma_e^2 > 0$ , sea  $\xi^{(1)} = \log(\sigma_e^2)$ ,  $\xi^{(1)}$  tiene soporte en  $\mathbb{R}$ . La función de densidad de  $\xi^{(1)}$  puede ser obtenida usando el método de transformación (Casella y Berger, 2002) y está dada por:

$$p(\xi^{(1)}|resto) \propto p(\sigma_e^2|resto) \exp \{ \xi^{(1)} \}.$$

En el algoritmo Metropolis con Caminata Aleatoria se genera  $\xi^{(1)}$  eligiendo un núcleo de transición propuesto para añadir ruido al estado actual. Suponiendo que el valor actual de  $\xi^{(1)}$  es  $\xi_{old}^{(1)}$  se desea actualizar su valor en la siguiente iteración a  $\xi_{new}^{(1)}$ , entonces se hace lo siguiente:

- a) Muestrear  $\xi^{(1)}$ ,  $\xi^{(1)} = \xi_{old}^{(1)} + \psi_1$ , donde  $\psi_1 \sim N(0, \eta_1^2)$ .
- b) Muestrear  $u$ ,  $U \sim U(0, 1)$ .
- c) Si  $u < \frac{p(\xi^{(1)}|resto)}{p(\xi_{old}^{(1)}|resto)}$  entonces  $\xi_{new}^{(1)} = \xi^{(1)}$ , de otro modo  $\xi_{new}^{(1)} = \xi_{old}^{(1)}$ .

Una vez que se ha obtenido  $\xi_{new}^{(1)}$ , calcular  $\sigma_e^{2,new} = \exp \{ \xi_{new}^{(1)} \}$ . El parámetro  $\eta_1^2$  se puede modificar para obtener una tasa de aceptación óptima.

#### 5. $p(\rho|resto)$

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

Finalmente, para  $\rho$ , vea que la distribución condicional tampoco tiene forma cerrada

$$p(\rho|resto) \propto \frac{S_U^n}{(1-\rho^2)^{\frac{n}{2}}} \exp \left\{ -\frac{S_U^2}{2(1-\rho^2)\sigma_e^2} \sum_{i=1}^n \left( \hat{e}_i + \frac{\sigma_e E_U}{S_U} - \frac{\sigma_e \rho z_i}{S_U} \right)^2 \right\} \\ \times \left( \frac{1-\rho}{2} \right)^{a_0-1} \left( 1 - \frac{1-\rho}{2} \right)^{b_0-1} I_{(-1,1)}(\rho).$$

Se propone utilizar el algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de  $\rho$ . Se utiliza la transformación de Fisher (1915) definida como  $\xi^{(2)} = \frac{1}{2} \log \left( \frac{1+\rho}{1-\rho} \right) = \tanh^{-1}(\rho)$ , el soporte de  $\xi^{(2)}$  está en  $\mathbb{R}$ . Además, se consideran los hiperparámetros  $a_0 = 1$  y  $b_0 = 1$  dando paso a una a priori  $U[0, 1]$ .

La función de  $\xi^{(2)}$  se puede obtener usando el método de transformación (Casella y Berger, 2002) y está dada por:

$$p(\xi^{(2)}|resto) \propto p(\rho|resto) \tanh(\xi^{(2)}).$$

En el algoritmo Metropolis con Caminata Aleatoria, se generó  $\xi^{(2)}$  eligiendo un núcleo de transición propuesto para añadir ruido al estado actual. Suponiendo que el valor actual de  $\xi^{(2)}$  es  $\xi_{old}^{(2)}$  se desea actualizar su valor en la siguiente iteración a  $\xi_{new}^{(2)}$ , entonces se hace lo siguiente:

- a) Muestrear  $\xi^{(2)}$ ,  $\xi^{(2)} = \xi_{old}^{(2)} + \psi_2$  donde  $\psi_2 \sim N(0, \eta_2^2)$ .
- b) Muestrear  $u$ ,  $U \sim U(0, 1)$ .
- c) Si  $u < \frac{p(\xi^{(2)}|resto)}{p(\xi_{old}^{(2)}|resto)}$  entonces  $\xi_{new}^{(2)} = \xi^{(2)}$ , de otro modo  $\xi_{new}^{(2)} = \xi_{old}^{(2)}$ .

Una vez que se ha obtenido  $\xi_{new}^{(2)}$ , calcular  $\rho^{new} = \tanh(\xi_{new}^{(2)})$ . El parámetro  $\eta_2^2$  se puede modificar para obtener una tasa de aceptación óptima.

Por lo anterior se tiene que hacer uso de diferentes técnicas computacionales para realizar las simulaciones, Muestreador de Gibbs para  $\beta_0$  y  $\beta$ , se utiliza el algoritmo de Metropolis con Caminata Aleatoria para lograr obtener muestras de  $\sigma_e^2$  (Kim, 2005), finalmente, para obtener muestras de  $\rho$ , se usa la transformación de Fisher (1915) y el algoritmo de Metropolis con Caminata Aleatoria.

Los pasos necesarios para el funcionamiento del muestreador de Gibbs y RWM para ajustar un modelo de regresión Bayesiana con errores aleatorios *Normal Asimétricos* (BSN), se describen en el Algoritmo 5. Dicho Algoritmo se implementó en el paquete estadístico *R* (R Core Team, 2018) y se presenta en el Anexo B.

## 2.10. Regresión con Errores aleatorios *Normal-Asimétricos*

**Algoritmo 5:** Regresión Bayesiana con errores aleatorios *SN* (*BSN*).

```

input : El algoritmo requiere que se le proporcione  $\mathbf{y}$ ,  $\mathbf{X}$ , y fijar  $R2$   $v_1$  y  $v_2$ .
/* donde  $v_1 \equiv \eta_1^2$  y  $v_2 \equiv \eta_2^2$  en el paso de Metropolis de  $\sigma^2$  y  $\rho$ . */
output:  $p(\boldsymbol{\theta}, \mathbf{z} | \text{resto})$ .

1 Obtener  $\boldsymbol{\theta}^{old} = (\beta_0^{old}, \beta_j^{old}, \sigma^{2,old}, \rho^{old})$ ,  $E_U$ ,  $V_U$ ,  $S_U$ ,  $z_i^{old}$  y fijar  $\sigma_\beta^2 = 1 \times 10^6$ .
2 for  $k$  in  $1 : B$  do /* con B número de iteraciones */
3   Obtener  $\hat{e}_i = y_i - (\beta_0^{old} + \mathbf{x}_i^t \boldsymbol{\beta}^{old})$ ,  $i = 1, \dots, n$ .
4   Muestrear de cada distribución posterior  $\boldsymbol{\theta} | \text{resto}$ , obtener  $\boldsymbol{\theta}^{new}$ , y actualizar
    $\hat{e}$  tras actualizar cada coeficiente de regresión  $\beta_0$  y  $\beta_j$ ,  $j = 1, \dots, p$ .
5   begin
6      $\beta_0 | \beta_j^{old}, \sigma^{2,old}, \rho^{old}, z_i^{old} \rightarrow \beta_0^{new}$ 
7      $\delta = \beta_0^{old} - \beta_0^{new}$  y  $\hat{e}_i = \hat{e}_i + \delta$ 
8     for  $j = 1, \dots, p$  do
9        $\beta_j | \beta_0^{new}, \sigma^{2,old}, \rho^{old}, z_i^{old} \rightarrow \beta_j^{new}$ 
10       $\delta = \beta_j^{old} - \beta_j^{new}$  y  $\hat{e}_i = \hat{e}_i + \delta x_{ij}$ 
11    end
12     $z_i | \beta_0^{new}, \beta_j^{new}, \sigma^{2,old}, \rho^{old} \rightarrow z_i^{new}$ 
13    Paso de Metropolis con Caminata Aleatoria para  $\sigma^2$  y  $\rho$ :
14    for  $h \in \{1, 2\}$  do
15      Muestrear de  $\xi^{(h)}$ ,  $\xi^{(h)} = \xi_{old}^{(h)} + \psi_h$ , con  $\psi_h \sim N(0, \eta_h^2)$ 
16      Muestrear de  $u$ ,  $U \sim U(0, 1)$ 
17      if  $u < \frac{p(\xi^{(h)} | \text{resto})}{p(\xi_{old}^{(h)} | \text{resto})}$  then
18         $\xi_{new}^{(h)} = \xi^{(h)}$ 
19      else
20         $\xi_{new}^{(h)} = \xi_{old}^{(h)}$ 
21      end
22      Para  $\sigma^2$  una vez que se ha obtenido  $\xi_{new}^{(1)}$ , calcular
       $\sigma^{2,new} = \exp \left\{ \xi_{new}^{(1)} \right\}$ . El parámetro  $\eta_1^2$  se puede modificar para
      obtener una tasa de aceptación óptima.
23      Para  $\rho$  una vez que se ha obtenido  $\xi_{new}^{(2)}$ , calcular  $\rho^{new} = \tanh \left( \xi_{new}^{(2)} \right)$ .
      El parámetro  $\eta_2^2$  se puede modificar para obtener una tasa de
      aceptación óptima.
24      Actualizar  $E_U$ ,  $V_U$  y  $S_U$ .
25    end
26  end
27 end

/* Las primeras  $m$  muestras son el calentamiento (burn-in) así que
se pueden desechar y tomar en cuenta las últimas  $B - m$  muestras
a las cuales se les aplica el adelgazamiento (thin). */

```

### 2.11. Distribución *t-Asimétrica* (*ST*)

La familia Normal Asimétrica es, por construcción, una clase de distribuciones más flexible que la Normal, debido al parámetro adicional  $\lambda$  que regula la asimetría. Este hecho, combinado con una serie de propiedades formales que disfruta esta distribución, lo convierte en una propuesta atractiva para el trabajo estadístico. Sin embargo, aunque es más flexible que la familia Normal, la presencia de un único parámetro para regular la forma de la densidad en (2.3) no puede ser suficiente para manejar adecuadamente los muy diversos tipos de situaciones que se cumplen en el trabajo práctico. Un problema frecuente es que las colas de la distribución observada son más gruesas que las normales, una situación que se encuentra en una gama de áreas de aplicación, casi sistemáticamente en econometría y finanzas.

Si bien el modelado estadístico escapa a reglas fijas, existe una percepción generalizada de que en un gran número de casos prácticos relacionados con el tipo continuo de observaciones con soporte ilimitado se puede producir un resultado satisfactorio trabajando con una distribución que permita una regulación independiente de localidad, escala, asimetría y curtosis.

En el marco de distribuciones asimétricas, la formulación actualmente más utilizada es proporcionada por la distribución *t-Asimétrica* (*ST*) (Arellano-Valle y Azzalini, 2013), destacando, entre esta familia, la proposición presentada por Azzalini y Capitanio (2003).

La familia *ST* ha sido considerada por primera vez por Branco y Dey (2001) pero en su forma multivariada, la función de densidad en la forma univariada ha sido dada por primera vez por Azzalini y Capitanio (2003) y por Gupta (2003).

A continuación se presentan diferentes alternativas para la distribución *t-Asimétrica* univariada propuestas por distintos autores, las cuales cuentan con diferentes características propias.

#### 2.11.1. *ST* Azzalini y Capitanio (2003), y Gupta (2003)

La distribución  $ST(\lambda, k)$  se puede obtener aplicando un factor de escala aleatorio adecuado a una variable aleatoria  $SN(\lambda)$ , replicando el mecanismo idéntico que conecta la distribución  $N(0, 1)$  y la distribución habitual de *t-Student*,  $t(k)$ . Esta representación estocástica para la distribución *ST* fue propuesta por Azzalini y Capitanio (2003) y consiste en lo siguiente:

Denote por  $W$  una variable aleatoria continua que puede ser obtenida como el cociente de otras dos variables aleatorias,

## 2.11. Distribución *t-Asimétrica* (*ST*)

---

$$W = \frac{U}{\sqrt{V/k}},$$

donde  $U \sim SN(\lambda)$  cuya función de densidad está dada por (2.3),  $V \sim \chi_k^2$  independiente de  $U$  y  $k$  son los grados de libertad de la variable  $V$ . Entonces, utilizando el método de Jacobianos (Casella y Berger, 2002) puede probarse que  $W \sim ST(\lambda, k)$ , donde  $\lambda \in \mathbb{R}$ , y  $k > 0$ .

Considere una transformación lineal de la variable aleatoria  $W$ ,

$$Y = \xi + \omega W$$

donde  $\xi \in \mathbb{R}$  es un parámetro de localidad y  $\omega > 0$  es un parámetro de escala. Utilizando el método de Jacobianos Casella y Berger (2002) puede obtenerse la distribución de  $Y$ , la cual está dada por:

$$p(y|\xi, \omega, \lambda, k) = \frac{2}{\omega} t(z; k) T\left(\lambda z \sqrt{\frac{k+1}{k+z^2}}; k+1\right), \quad z = \frac{y-\xi}{\omega} \quad (2.8)$$

donde  $y \in \mathbb{R}$ ,  $t(\cdot; k)$  y  $T(\cdot; k)$  denotan la función de densidad de probabilidades y función de distribución acumulada, respectivamente, de una variable *t-Student* con  $k > 0$  grados de libertad,  $\xi \in \mathbb{R}$ ,  $\omega > 0$  y  $\lambda \in \mathbb{R}$  parámetro de sesgo. Si  $Y$  tiene la función de densidad dada en (2.8), entonces  $Y \sim ST(\xi, \omega^2, \lambda, \nu)$ , (Azzalini y Capitanio, 2003). Algunas propiedades de la distribución *ST* presentada de esta forma son:

- 1) Si  $k \rightarrow \infty$  en (2.8) entonces  $Y \sim SN_D(\lambda)$ .
- 2) Si  $k = 1$  se obtiene la distribución Skew-Cauchy,  $SC(\xi, \omega^2, \lambda)$ .
- 3) Si  $\lambda = 0$  entonces  $Y \sim t(\xi, \omega^2, k)$ .
- 4) Si  $k \rightarrow \infty$  y  $\lambda = 0$  la distribución  $ST(\xi, \omega^2, \lambda, k)$  converge a la distribución  $N(\xi, \omega^2)$ .
- 5) Si  $Y \sim ST(\lambda, k)$ , entonces  $-Y \sim ST(-\lambda, k)$ .
- 6) Si  $Y \sim ST(\lambda, k)$ , entonces  $Y^2 \sim F(1, k)$  (Di Ciccio y Monti, 2011).

Un aspecto problemático con este tipo de formulación es que los parámetros no están relacionados con los momentos de una manera simple. Para ejemplificar el hecho, el parámetro de localidad  $\xi$  no corresponde a ninguna cantidad utilizada tradicionalmente

## 2.11. Distribución *t-Asimétrica* (*ST*)

para cuantificar la ubicación, como la media o la mediana, ni  $\xi$  está relacionado con ellas en una forma simple, un hecho que afecta la interpretación de este parámetro. Con otros parámetros, especialmente aquellos que reflejan la asimetría y la curtosis, el problema no es más sencillo, y se vuelve aún más complejo en el caso multivariado. Una complicación adicional proviene de la falta de momentos de orden no mayores que  $k$ . Sin embargo, hay que tener en consideración que estos problemas no son específicos de la formulación *ST*, y también presentan diversos grados de complejidad en propuestas alternativas ([Arellano-Valle y Azzalini, 2013](#)).

A diferencia del caso de la *Normal Asimétrica*, en la *ST* de [Azzalini y Capitanio \(2003\)](#) y [Gupta \(2003\)](#) no se tiene el problema de singularidad y la llamada parametrización centrada, a saber,  $\mu$ ,  $\sigma^2$  y  $\lambda$ , que representan la media, la varianza y el índice de asimetría, se usa solo con fines de interpretación.

En la Figura 2.3 se puede observar la densidad de la *ST* definida por [Azzalini y Capitanio \(2003\)](#) y [Gupta \(2003\)](#), la cual tiene la forma presentada en (2.8), para distintos valores de  $\lambda$ , parámetro de sesgo, y de  $k$ , parámetro de forma. Algunas de las propiedades enunciadas se pueden identificar con facilidad en la Figura 2.3.

De la Figura 2.3 vea que el gráfico a) es idéntico que la Figura 2.2 de la distribución *Normal Asimétrica*, como ya se mencionó, esto se debe a que cuando  $k \rightarrow \infty$  la distribución *ST* converge a la *SN*.

### 2.11.2. *ST* Zhu y Galbraith (2010)

En general, dentro de la literatura, se describen varios tipos de distribuciones *ST* con un parámetro que controla principalmente la asimetría de la parte central de una distribución. Por lo tanto, una clase de distribuciones asimétricas que tiene un parámetro de asimetría y dos parámetros de cola ofrece la posibilidad de mejorar nuestra capacidad para ajustar y pronosticar datos empíricos en las regiones finales que son críticas, por ejemplo, en problemas de gestión de riesgos y otras aplicaciones econométricas financieras.

Bajo este contexto [Zhu y Galbraith \(2010\)](#) proponen una distribución *ST* cuya función de densidad está dada por:

$$p(y|\mu, \sigma, \alpha, k_1, k_2) = \begin{cases} \frac{B}{\sigma} \left[ 1 + \frac{B^2}{k_1} \left( \frac{y - \mu}{2\sigma\alpha Q(k_1)} \right)^2 \right]^{-\frac{k_1+1}{2}} & \text{si } y \leq \mu \\ \frac{B}{\sigma} \left[ 1 + \frac{B^2}{k_2} \left( \frac{y - \mu}{2\sigma(1-\alpha)Q(k_2)} \right)^2 \right]^{-\frac{k_2+1}{2}} & \text{si } y > \mu \end{cases} \quad (2.9)$$

## 2.11. Distribución *t-Asimétrica (ST)*

---

donde  $y \in \mathbb{R}$ ,  $Q(k) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)\sqrt{\pi k}}$ ,  $B \equiv \alpha Q(k_1) + (1-\alpha)Q(k_2)$  y  $\Gamma(\cdot)$  es la función gamma;  $\mu \in \mathbb{R}$  parámetro de localidad,  $\sigma > 0$  parámetro de escala,  $\alpha \in (0, 1)$  parámetro de sesgo, y dos parámetros de forma  $k_1 > 0$  y  $k_2 > 0$  correspondientes a las colas.

Una representación estocástica de la *ST* definida de esta manera que resulta útil para estudiar las propiedades de la distribución y en estudios de simulación se da a continuación.

Denote por  $T(k)$  una variable aleatoria que tiene la distribución *t-Student* estándar con  $k$  grados de libertad. Considere tres variables aleatorias independientes  $U \sim U(0, 1)$ ,  $T(k_1)$  y  $T(k_2)$ .

$$Y = \alpha^* |T(k_1)| [\text{sign}(U - \alpha) - 1] + (1 - \alpha^*) |T(k_2)| [\text{sign}(U - \alpha) + 1] \quad (2.10)$$

donde  $\alpha^* = \frac{\alpha Q(k_1)}{\alpha Q(k_1) + (1 - \alpha) Q(k_2)}$ ,  $\text{sign}(x) = +1$  si  $x > 0$ ,  $-1$  si  $x < 0$ , y  $0$  si  $x = 0$ .

Entonces, es posible mostrar que  $Y$  es una variable aleatoria *ST* estándar que tiene la densidad con los tres parámetros de forma  $(\alpha, k_1, k_2)$ .

[Naranjo et al. \(2017\)](#) propone otra representación estocástica de la *ST* definida en (2.9) por medio de una mezcla de distribuciones *Normales*, la cual parece ser adecuada en estudios de simulación ya que es posible obtener distribuciones condicionales de los parámetros de interés.

$$f(y) = \alpha \int f(y|\gamma_1) f(\gamma_1) d\gamma_1 + (1 - \alpha) \int f(y|\gamma_2) f(\gamma_2) d\gamma_2 \quad (2.11)$$

donde

$$Y|\gamma_1 \sim N(\mu, [2\sigma\alpha Q(k_1)]^2 \gamma_1^{-1}) I(Y \leq \mu), \quad \gamma_1 \sim Ga(k_1/2, \text{scale} = 2/k_1)$$

$$Y|\gamma_2 \sim N(\mu, [2\sigma(1-\alpha)Q(k_2)]^2 \gamma_2^{-1}) I(Y > \mu), \quad \gamma_2 \sim Ga(k_2/2, \text{scale} = 2/k_2)$$

con  $\alpha \in (0, 1)$ , entonces  $f(y)$  es la fdp de una distribución *ST*, es decir,  $Y \sim ST(\mu, \sigma^2, \alpha, k_1, k_2)$ .

## 2.11. Distribución *t-Asimétrica* (*ST*)

---

En la Figura 2.4 se puede observar la densidad de la *ST* definida por [Zhu y Galbraith \(2010\)](#), la cual tiene la forma presentada en (2.9), para distintos valores de  $\alpha$ , parámetro de sesgo, y de  $k_1$ , parámetro de forma en la cola izquierda. Algunas de las propiedades enunciadas se pueden identificar con facilidad en la Figura 2.4.

### 2.11.3. *ST* [Fernández y Steel \(1998\)](#)

[Fernández y Steel \(1998\)](#) proponen una metodología que permite transformar distribuciones simétricas en asimétricas de una manera muy sencilla, la misma solamente requiere usar un escalar  $\lambda$ , es decir, un parámetro de asimetría, para hacer dicha transformación. Particularmente usan esta metodología para transformar una distribución de densidad *t-Student* en una distribución *t-Asimétrica*.

Primeramente se considera  $Y \sim t(k)$ , donde  $k > 2$ . La metodología propuesta introduce asimetría en las funciones de densidad mediante factores de escala inversos en los valores positivos y negativos de  $Y$ . Estos factores se definen con el escalar  $\lambda > 0$ . Particularmente, si este escalar es fijo, la función de densidad de la variable  $Y$  que sigue una distribución *t-Asimétrica* queda definida como sigue:

$$p(y|\lambda, k) = \frac{2}{(\lambda + 1/\lambda)} \left[ t\left(\frac{y}{\lambda}; k\right) I_{[0, \infty)}(y) + t(\lambda y; k) I_{(-\infty, 0)}(y) \right], \quad (2.12)$$

donde  $t(\cdot, k)$  denota la fdp de una variable aleatoria *t-Student* con  $k > 2$  grados de libertad. La fdp anterior generaliza la distribución *t-Student* con base en el parámetro de asimetría  $\lambda$ . Si  $\lambda = 1$ , la función resultante es la distribución *t-Student* simétrica. Si  $\lambda \neq 1$ , la función resultante es una *ST*. Particularmente, si  $\lambda < 1$ , se tiene una distribución sesgada a la izquierda, en contraparte, si  $\lambda > 1$ , se tiene una distribución sesgada a la derecha. Entonces, si la variable  $Y$  se distribuye siguiendo una *ST* estandarizada puede denotarse como  $Y \sim ST(\lambda, k)$  ([Lorenzo y Ruiz, 2014](#)).

En la Figura 2.5 se puede observar la densidad de la *ST* definida por [Fernández y Steel \(1998\)](#), la cual tiene la forma presentada en (2.12), para distintos valores de  $\lambda$ , parámetro de sesgo, y de  $k$ , parámetro de forma. Algunas de las propiedades enunciadas se pueden identificar con facilidad en la Figura 2.5.

### 2.11.4. *ST* [Jones y Faddy \(2003\)](#)

[Jones y Faddy \(2003\)](#) proponen una familia de distribuciones que incluye la distribución *t-Student* como caso especial, y también incluye extensiones aún tomando valores en toda la línea real con asimetría distinta de cero.

## 2.11. Distribución *t-Asimétrica (ST)*

---

$$p(y|a, b) = C_{a,b}^{-1} \left\{ 1 + \frac{y}{(a+b+y^2)^{1/2}} \right\}^{a+1/2} \left\{ 1 - \frac{y}{(a+b+y^2)^{1/2}} \right\}^{b+1/2}, \quad y \in \mathbb{R} \quad (2.13)$$

donde

$$C_{a,b} = 2^{a+b-1} B(a, b) (a+b)^{1/2},$$

con  $a, b \in \mathbb{R}$  y  $B(\cdot)$  denota la función beta,  $B(x, y) = \int_0^1 u^{x-1} (1-u)^{y-1} du$ ,  $x, y \in \mathbb{R}^+$ . Cuando  $a = b$ ,  $p(\cdot)$  se reduce a la distribución *t-Student* con  $2a$  grados de libertad. Vea que si  $a, b \rightarrow \infty$   $p(\cdot)$  converge a la distribución  $N(0, 1)$ . Cuando  $a < b$  o  $a > b$ ,  $p(\cdot)$  está negativa o positivamente sesgada respectivamente, de hecho  $p(y|b, a) = p(-y|a, b)$ .

Se plantea una transformación de variables aleatorias para obtener la distribución en (2.13) la cual consiste en partir de una v.a.  $Y \sim \text{Beta}(a, b)$ , con  $a = 0$  y  $b = 1$ , entonces considere la siguiente transformación:

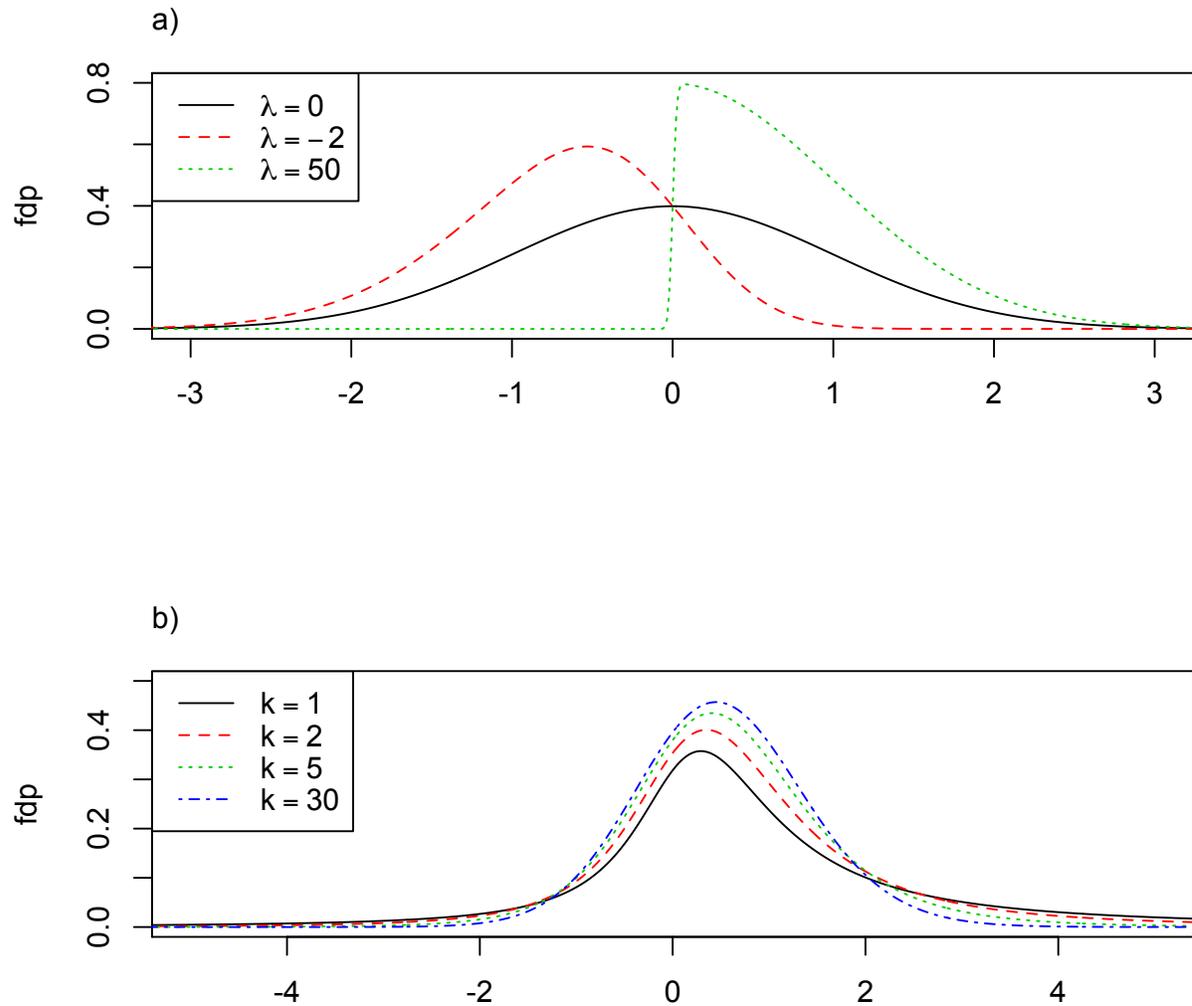
$$T = \frac{\sqrt{(a+b)(2Y-1)}}{2\sqrt{Y(1-Y)}}.$$

Una formulación equivalente está dada por la relación entre una variable aleatoria con distribución Beta y un par de variables aleatorias distribuidas como Chi-cuadrada. Sean  $U \sim \chi_{2a}^2$  y  $V \sim \chi_{2b}^2$  variables aleatorias independientes entonces:

$$T = \frac{\sqrt{(a+b)(U-V)}}{2\sqrt{UV}}.$$

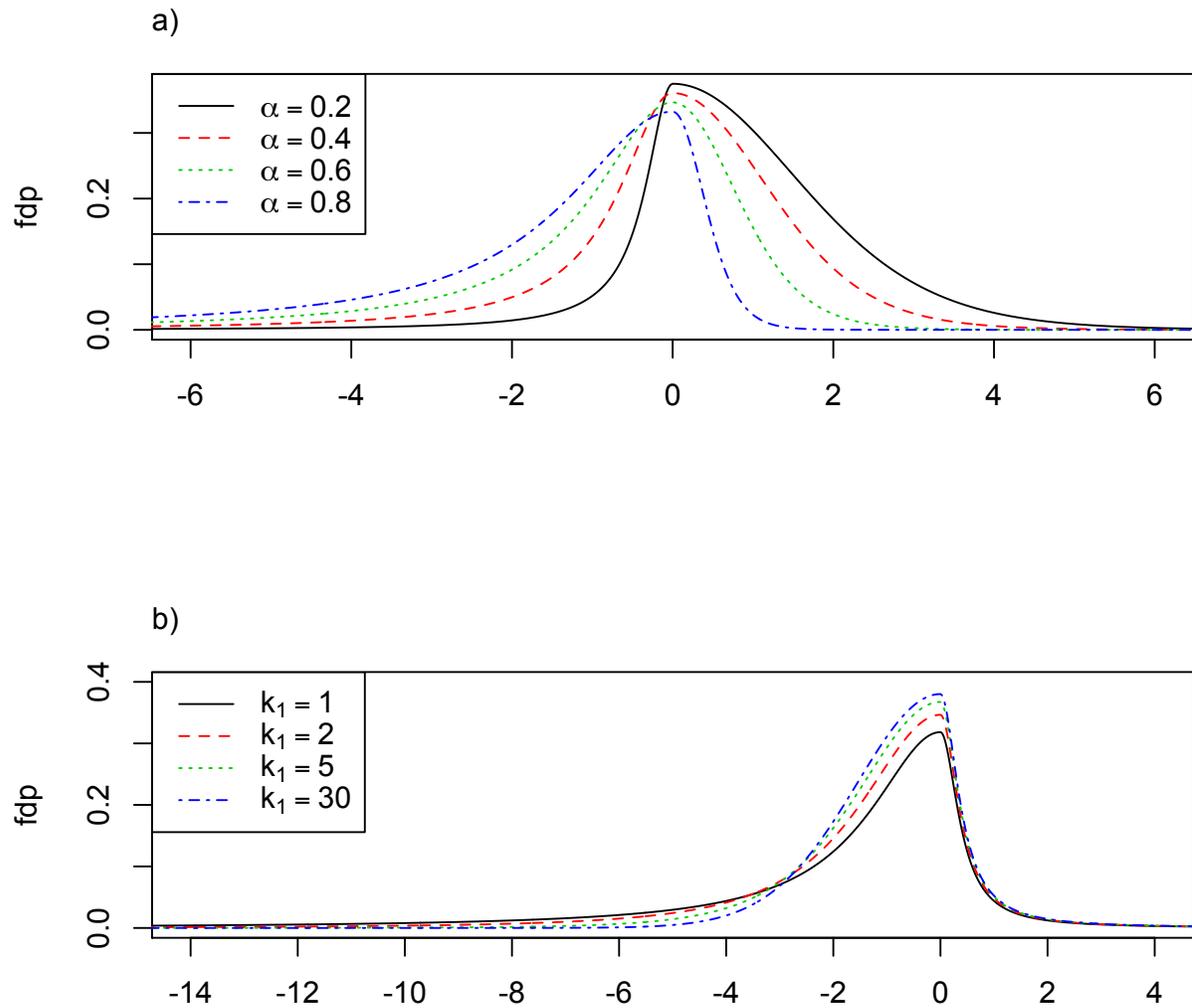
Finalmente, otra representación para esta última ecuación está dada por  $T = \frac{1}{2} \sqrt{(a+b)(W^{1/2} - W^{-1/2})}$  donde  $W = aF/b$  con  $F \sim F_{2a, 2b}$ .

## 2.11. Distribución $t$ -Asimétrica ( $ST$ )



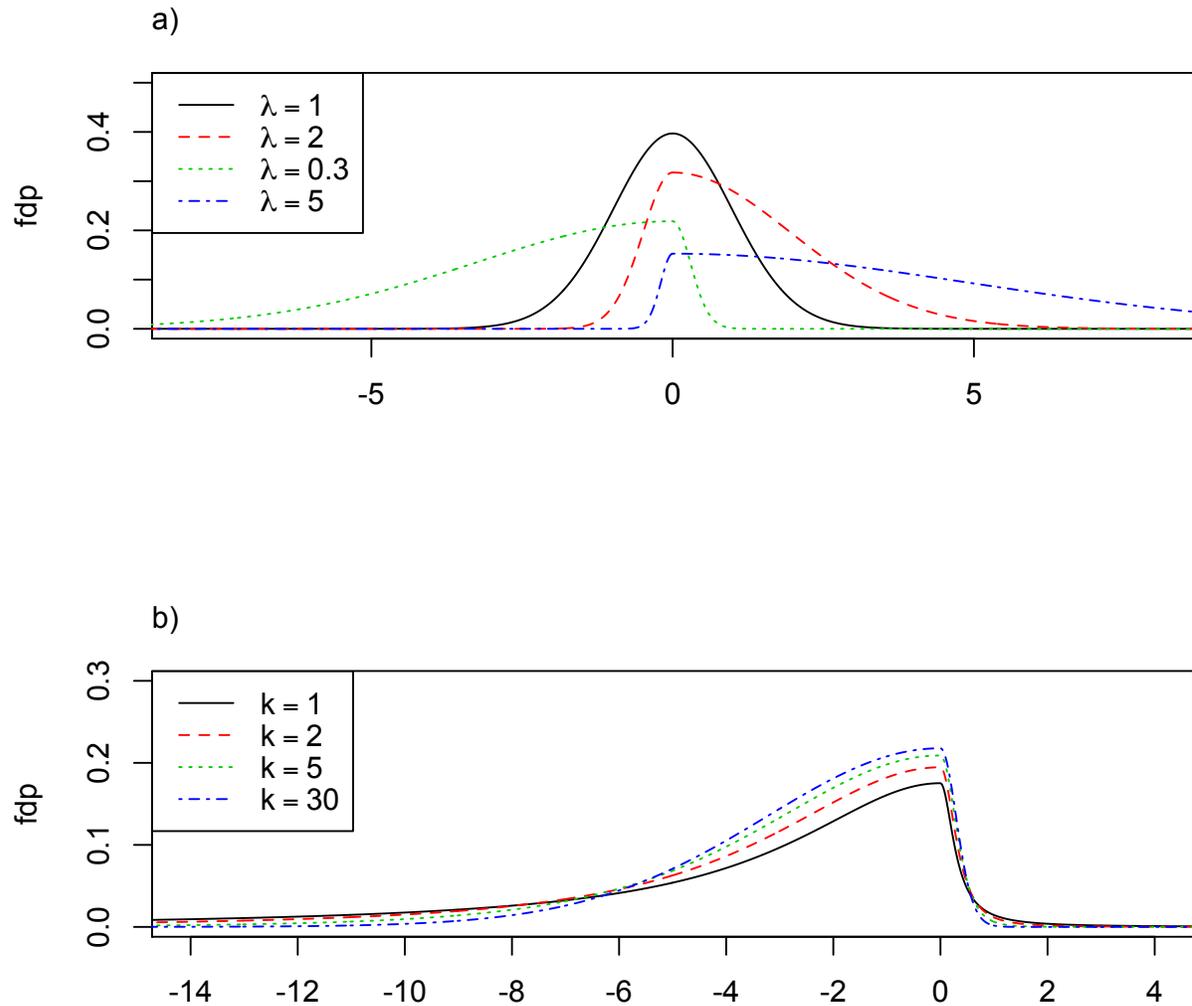
**Figura 2.3:** Densidad de la  $ST$  ( $\xi, \omega^2, \lambda, k$ ) definida por [Azzalini y Capitanio \(2003\)](#) y [Gupta \(2003\)](#) con distintos parámetros de sesgo,  $\lambda$ , y de forma,  $k$ .  
a) Densidad  $ST$  ( $\xi = 0, \omega^2 = 1, \lambda, k = \infty$ ) con distintos parámetros de sesgo,  $\lambda = \{0, -2, 50\}$ . Observe que cuando  $\lambda \neq 0$  y  $k \rightarrow \infty$  la distribución  $ST$  converge a la  $SN$ . Además, cuando  $\lambda = 0$  y  $k \rightarrow \infty$  la distribución  $ST$  converge a la distribución  $N(\xi, \omega^2)$ .  
b) Densidad  $ST$  ( $\xi = 0, \omega^2 = 1, \lambda = 0.8, k$ ) con distintos parámetros de forma,  $k = \{1, 2, 5, 30\}$ .

## 2.11. Distribución $t$ -Asimétrica ( $ST$ )



**Figura 2.4:** Densidad de la  $ST$  ( $\mu, \sigma^2, \alpha, k_1, k_2$ ) definida por [Zhu y Galbraith \(2010\)](#) con distintos parámetros de sesgo,  $\alpha$ , y de forma en la cola izquierda,  $k_1$ .  
a) Densidad  $ST$  ( $\mu = 0, \sigma^2 = 1, \alpha, k_1 = 1, k_2 = 10$ ) con distintos parámetros de sesgo,  $\alpha = \{0.2, 0.4, 0.6, 0.8\}$ .  
b) Densidad  $ST$  ( $\mu = 0, \sigma^2 = 1, \alpha = 0.8, k_1, k_2 = 1$ ) con distintos parámetros de forma en la cola izquierda,  $k_1 = \{1, 2, 5, 30\}$ .

## 2.11. Distribución $t$ -Asimétrica ( $ST$ )



**Figura 2.5:** Densidad de la  $ST(\lambda, k)$  definida por Fernández y Steel (1998) con distintos parámetros de sesgo,  $\lambda$ , y de forma,  $k$ .  
a)  $ST(\lambda, k = \infty)$ , con distintos parámetros de sesgo,  $\lambda = \{1, 2, 0.3, 5\}$ . Observe que cuando  $\lambda = 1$  se obtiene la distribución  $t(k)$ .  
b)  $ST(\lambda = 0.8, k)$  con distintos parámetros de forma,  $k = \{1, 2, 5, 30\}$ .

## 2.12. Elección de una adecuada Distribución $ST$

La elección de una representación de la distribución  $ST$  dependerá de las preferencias y necesidades particulares de cada usuario de esta. Existen más propuestas de las que aquí han sido presentadas para el caso univariado, y lo mismo ocurre para el caso multivariado. Cabe destacar que algunas de estas proposiciones son equivalentes entre ellas bajo alguna reparametrización.

Así, la elección puede reducirse a criterios como la facilidad de aplicación de la representación de la distribución  $ST$ , el número de parámetros que esta utiliza, la tratabilidad matemática o el mayor uso y aplicación en diferentes trabajos, por ejemplo.

En atención a lo anterior, en este trabajo se ha optado por desarrollar el problema de regresión con dos proposiciones de las presentadas anteriormente, la distribución  $ST$  de [Azzalini y Capitanio \(2003\)](#) y la  $ST$  de [Zhu y Galbraith \(2010\)](#). En cuanto a la primera, se trata de una de las formulaciones más utilizadas en distintos trabajos de investigación, presenta una representación estocástica intuitiva, es matemáticamente tratable para el desarrollo de los resultados Bayesianos de interés para esta investigación y presenta un número de parámetros menor comparada con otras proposiciones. En cuanto a la segunda, a pesar de tener un número mayor de parámetros, su autor manifiesta que esto permite manejar la forma de las colas independientemente, lo que podría ser útil en ocasiones, además es relativamente reciente y poco estudiada por lo que resulta interesante abordarla.

De esta manera, en el Capítulo 3 se presentan los desarrollos teórico-matemáticos que permiten la estimación de los parámetros, bajo el contexto de regresión, de las dos distribuciones  $ST$  anteriores y desde un enfoque Bayesiano. Cabe hacer notar que para el caso de la formulación de [Zhu y Galbraith \(2010\)](#) el trabajo culmina con la exposición de los desarrollos realizados, mientras que para la proposición de [Azzalini y Capitanio \(2003\)](#) se ha hecho además un programa, utilizando los desarrollos expuestos en el capítulo siguiente, en el paquete estadístico  $R$  ([R Core Team, 2018](#)) que permite el ajuste de modelos de RLM, el cual es utilizado posteriormente.

# Capítulo 3

## Regresión con Errores Aleatorios *ST*

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

Si se considera el modelo de regresión lineal múltiple

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\ e_i \sim iidST(0, \omega^2, \lambda, k),$$

donde  $ST(0, \omega^2, \lambda, k)$  denota una variable aleatoria con función de densidad dada por (2.8).

Utilizando la representación estocástica propuesta por [Azzalini y Capitanio \(2003\)](#) y desarrollada a detalle por [Parisi y Liseo \(2018\)](#) puede obtenerse la distribución  $ST(0, 1, \lambda, k)$  como sigue:

Denote por  $W$  una variable continua que puede ser obtenida como el cociente de otras dos variables aleatorias

$$W = \frac{U}{\sqrt{V/k}},$$

donde  $U \sim SN(\lambda)$ ,  $V \sim \chi_k^2$  independiente de  $U$  y  $k$  son los grados de libertad de la variable  $V$ .

Considere la distribución  $p(u, z) = p(u|z, \rho) p(z)$ , con  $p(u|z, \rho)$  la función de densidad de una variable aleatoria *Normal* con media  $\rho z$  y varianza  $1 - \rho^2$  y  $p(z)$  la función

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

de densidad de una variable aleatoria *Normal Truncada*, con parámetro de localidad 0, parámetro de escala 1, cota inferior de truncamiento 0 y cota superior de truncamiento  $\infty$ . La distribución conjunta de  $U$  y  $Z$  está dada por (2.5) y a partir de esta es posible obtener una variable aleatoria *SN* usando aumentación de datos.

Obteniendo la densidad conjunta de  $U$ ,  $Z$  y  $V$ :

$$\begin{aligned}
 p(u, z, v|\rho, k) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u-\rho z)^2\right\} \\
 &\times \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} I_z(0, \infty) \times \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} v^{\frac{k}{2}-1} \exp\left\{-\frac{v}{2}\right\},
 \end{aligned} \tag{3.1}$$

donde  $u \in \mathbb{R}$ ,  $z > 0$ ,  $v > 0$ ,  $\rho \in (-1, 1)$  y  $k > 0$ .

Note que  $p(u|\rho)$  se obtiene mediante integración de  $p(u, z, v|\rho, k)$  dada en (3.1), es decir,  $p(u|\rho) = \int \int p(u, z, v|\rho, k) dv dz$ , obteniéndose  $U \sim SN(\lambda)$ , con  $\lambda = \frac{\rho}{\sqrt{1-\rho^2}}$ .

Defina la transformación  $W = \frac{U}{\sqrt{V/k}}$  y obtenga su distribución a partir de  $p(u, z, v|\rho, k)$  dada en 3.1 utilizando el método de Jacobianos ([Casella y Berger, 2002](#)). Esto es, obtenga la distribución conjunta  $p(w, z, v|\rho, k)$  y posteriormente, al resolver la integral doble  $\int \int p(w, z, v|\rho, k) dv dz = p(w|\rho, k)$ , se demuestra que  $W \sim ST(0, 1, \lambda, k)$  con  $\lambda = \frac{\rho}{\sqrt{1-\rho^2}}$ .

Ahora considere una transformación lineal de la variable aleatoria  $W$ ,

$$Y = \xi + \omega W,$$

donde  $\xi \in \mathbb{R}$  es un parámetro de localidad y  $\omega > 0$  es un parámetro de escala. De nueva cuenta, utilizando el método de Jacobianos ([Casella y Berger, 2002](#)) puede obtenerse la distribución conjunta de  $Y$ ,  $Z$  y  $V$ , la cual está dada por:

$$\begin{aligned}
 p(y, z, v|\xi, \omega, \rho, k) &= \frac{1}{\omega} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\xi}{\omega}\right)\sqrt{\frac{v}{k}} - \rho z\right]^2\right\} \\
 &\times \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} I_z(0, \infty) \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} v^{\frac{k}{2}-1} \exp\left\{-\frac{v}{2}\right\} \sqrt{\frac{v}{k}},
 \end{aligned} \tag{3.2}$$

donde  $y \in \mathbb{R}$ ,  $z > 0$ ,  $v > 0$ ,  $\xi \in \mathbb{R}$ ,  $\omega > 0$ ,  $\rho \in (-1, 1)$  y  $k > 0$ .

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

Entonces resolviendo la integral doble

$$p(y|\xi, \omega, \rho, k) = \int_0^\infty \int_0^\infty p(y, z, v|\xi, \omega, \rho, k) dv dz$$

es posible mostrar que  $Y \sim ST(\xi, \omega^2, \lambda, k)$ , es decir,  $Y$  tiene la forma de una función de densidad dada en (2.8), con  $\lambda = \frac{\rho}{\sqrt{1-\rho^2}}$ . Esta representación es muy conveniente porque nos permite escribir una función de probabilidad aumentada, como si hubiésemos observado los valores latentes  $z$  y  $v$ .

Entonces, siguiendo en el contexto del problema de regresión lineal múltiple planteado al comienzo de esta sección, donde  $e_i \sim iidST(0, \omega^2, \lambda, k)$ , por propiedades de la distribución *t-Asimétrica* se tiene que

$$Y_i \sim ST(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \omega^2, \lambda, k), \text{ independientes.}$$

Sea  $\xi_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}$ . La verosimilitud aumentada está dada por:

$$\begin{aligned} \prod_{i=1}^n p(y_i, z_i, v_i | \boldsymbol{\theta}) &= \prod_{i=1}^n \frac{1}{\omega} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{y_i - \xi_i}{\omega} \right) \sqrt{\frac{v_i}{k}} - \rho z_i \right]^2 \right\} \\ &\times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} z_i^2 \right\} I_{z_i}(0, \infty) \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} v_i^{\frac{k}{2}-1} \exp \left\{ -\frac{v_i}{2} \right\} \sqrt{\frac{v_i}{k}}. \end{aligned}$$

Note que el vector de parámetros  $\boldsymbol{\theta}$  incluye los coeficientes de regresión  $\beta_0$  y  $\boldsymbol{\beta}$  y los parámetros  $\omega^2$ ,  $\rho$  y  $k$ , es decir,  $\boldsymbol{\theta} = [\beta_0, \boldsymbol{\beta}, \omega^2, \rho, k]^t$  con  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t$ , mientras que alternativamente  $\mathbf{z} = [z_1, z_2, \dots, z_n]^t$  y  $\mathbf{v} = [v_1, v_2, \dots, v_n]^t$  son considerados pesos. Esto se conoce como “Aumentación de datos” ([Tanner y Wong, 1987](#)).

Con el fin de especificar el modelo completo en el marco Bayesiano se asignan las siguientes distribuciones a priori:

$$\begin{aligned} \beta_0 | \sigma_\beta^2 &\sim N(0, \sigma_\beta^2), \\ \boldsymbol{\beta} | \sigma_\beta^2 &\sim NM(\mathbf{0}, \sigma_\beta^2 \mathbf{I}), \\ p(\omega^2) &\propto \frac{1}{\omega^2}. \end{aligned}$$

Cabe señalar que para la distribución a priori de  $\beta_0$  y  $\boldsymbol{\beta}$  se fijó  $\sigma_\beta^2 = 1 \times 10^6$  dando lugar a una distribución a priori no informativa.

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

Para el parámetro  $\rho$ , basados en la transformación  $\rho = 1 - 2B$ , donde  $B \sim \text{Beta}(a_0, b_0)$ , la densidad a priori queda definida como:

$$p(\rho|a_0, b_0) \propto \left(\frac{1-\rho}{2}\right)^{a_0-1} \left(1 - \frac{1-\rho}{2}\right)^{b_0-1} I_{(-1,1)}(\rho).$$

Finalmente, el parámetro  $k$  es considerado como fijo debido a que se tienen problemas de convergencia del algoritmo al momento de probar diferentes distribuciones a priori para este parámetro.

Con el fin de obtener estimaciones de  $\boldsymbol{\theta}$  se aplica el Teorema Bayes utilizando la función de verosimilitud de los datos completos, la distribución a priori de  $\boldsymbol{\theta}$  y  $\boldsymbol{\Omega}$  el conjunto de hiperparámetros, el muestreador de Gibbs y el algoritmo de Metropolis con Caminata Aleatoria.

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{v}|\text{datos}) &\propto p(\mathbf{y}, \mathbf{z}, \mathbf{v}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\Omega}) \\ &= \prod_{i=1}^n [p(y_i, z_i, v_i|\boldsymbol{\theta})] p(\beta_0|\sigma_\beta^2) p(\boldsymbol{\beta}|\sigma_\beta^2) p(\omega^2) p(\rho|a_0, b_0). \end{aligned}$$

Entonces utilizando la función de verosimilitud de los datos aumentados y las densidades a priori antes definidas, la distribución a posteriori  $p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{v}|\text{datos})$  está dada por:

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{v}|\text{datos}) &\propto p(\mathbf{y}, \mathbf{z}, \mathbf{v}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\Omega}) \\ &= \exp \left\{ -\frac{1}{2(1-\rho^2)\omega^2} \sum_{i=1}^n \left[ (y_i - \xi_i) \sqrt{\frac{v_i}{k}} - \rho\omega z_i \right]^2 \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n z_i^2 \right\} \\ &\quad \times \left(\frac{1}{\omega}\right)^n \left(\frac{1}{\sqrt{1-\rho^2}}\right)^n \left(\frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}\right)^n \prod_{i=1}^n \left[ v_i^{\frac{k}{2}-1} \sqrt{\frac{v_i}{k}} \right] \exp \left\{ -\frac{1}{2} \sum_{i=1}^n v_i \right\} \\ &\quad \times N(\beta_0|0, \sigma_\beta^2) N(\boldsymbol{\beta}|\mathbf{0}, \sigma_\beta^2 \mathbf{I}) \frac{1}{\omega^2} \left(\frac{1-\rho}{2}\right)^{a_0-1} \left(1 - \frac{1-\rho}{2}\right)^{b_0-1}. \end{aligned} \tag{3.3}$$

Sea  $\hat{e}_i = y_i - \xi_i$ . Entonces las distribuciones condicionales completas necesarias para implementar el algoritmo del muestreador de Gibbs y RWM se pueden obtener de (3.3) y se presentan a continuación.

3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

1.  $p(\beta_0|resto)$

$$\beta_0|resto \sim N\left(\frac{c_2}{c_1}, \frac{(1-\rho^2)\omega^2}{c_1}\right),$$

donde

$$c_1 = \sum_{i=1}^n \left(\frac{v_i}{k}\right) + \frac{(1-\rho^2)\omega^2}{\sigma_\beta^2}, \quad c_2 = \sum_{i=1}^n \left(d_i \sqrt{\frac{v_i}{k}}\right),$$

$$d_i = y_i^* \sqrt{\frac{v_i}{k}} - \rho\omega z_i, \quad y_i^* = \hat{e}_i + \beta_0.$$

2.  $p(\beta_j|resto), j = 1, \dots, p$

$$\beta_j|resto \sim N\left(\frac{c_2}{c_1}, \frac{(1-\rho^2)\omega^2}{c_1}\right),$$

donde

$$c_1 = \sum_{i=1}^n \left(\frac{x_{ij}^2 v_i}{k}\right) + \frac{(1-\rho^2)\omega^2}{\sigma_\beta^2}, \quad c_2 = \sum_{i=1}^n \left(f_i x_{ij} \sqrt{\frac{v_i}{k}}\right),$$

$$f_i = y_i^* \sqrt{\frac{v_i}{k}} - \rho\omega z_i, \quad y_i^* = \hat{e}_i + \beta_j x_{ij}.$$

3.  $p(z_i|resto), i = 1, \dots, n$

$$z_i|resto \sim NT\left(\mu = g_i \frac{\rho}{\omega}, \sigma^2 = 1 - \rho^2, a = 0, b = \infty\right),$$

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

donde  $g_i = \hat{e}_i \sqrt{\frac{v_i}{k}}$ .

#### 4. $p(v_i|resto)$

Para la variable latente  $V_i$  no es posible identificar el núcleo de alguna distribución univariada conocida, pues su distribución condicional posterior no tiene forma cerrada

$$p(v_i|resto) \propto v_i^{C-1} \exp\{-A_i v_i - B_i \sqrt{v_i}\},$$

donde

$$A_i = \frac{b_i^2 + (1 - \rho^2) \omega^2}{2(1 - \rho^2) \omega^2}, \quad B_i = -\frac{2b_i \rho \omega z_i}{(1 - \rho^2) \omega^2}, \quad C = \frac{k + 1}{2}, \quad b_i = \frac{\hat{e}_i}{\sqrt{k}}.$$

[Parisi y Liseo \(2018\)](#) proponen usar el Algoritmo de Aceptación y Rechazo ([Gelman \*et al.\*, 2014](#)) para muestrear de  $V_i|resto$ . Teniendo la distribución condicional completa  $p(v_i|resto)$  como densidad objetivo se hace lo siguiente:

- a) Definir la Distribución Generadora de Candidatos  $G = R^2$ , con  $R \sim \Gamma(\alpha_r, \gamma_r)$ , entonces usando Jacobianos se encuentra la distribución de  $G$ , la cual esta dada por:

$$p_G(g|\alpha_r, \gamma_r) = \frac{\gamma_r^{\alpha_r}}{2\Gamma(\alpha_r)} g^{\frac{\alpha_r}{2}-1} \exp\{-\gamma_r \sqrt{g}\}, \quad g > 0.$$

- b) Definir

$$\alpha_r^* = 2C = k + 1,$$

$$\gamma_r^* = \frac{1}{2} \left( B_i + \sqrt{B_i^2 + 8A_i(2C + 1)} \right),$$

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

usando los parámetros  $\alpha_r^*$  y  $\gamma_r^*$  se optimiza la eficiencia del algoritmo de Aceptación y Rechazo, entonces se tiene  $p_G(g|\alpha_r^*, \gamma_r^*)$ .

c) El algoritmo de Aceptación y Rechazo requiere la obtención de la constante  $M$

- Paso de Rechazo:

$$p(v_i|resto) \leq Mp_G(v_i; \alpha_r^*, \gamma_r^*).$$

Razón de Importancia:

$$\frac{p(v_i|resto)}{p_G(v_i; \alpha_r^*, \gamma_r^*)} = m(v_i),$$

dados los parámetros de la distribución generadora de candidatos, esta función tiene un máximo en

$$v_i^* = \left( \frac{\gamma_r^* - B_i}{2A_i} \right)^2,$$

así el valor de  $M$  puede ser obtenido finalmente como

$$M = m(v_i^*).$$

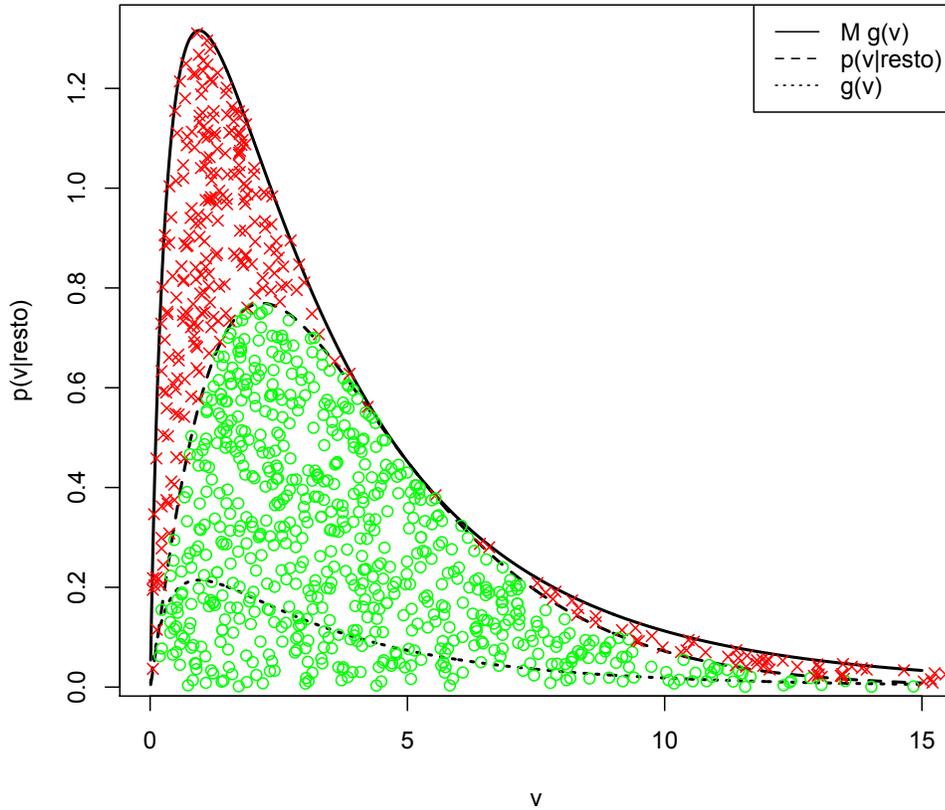
d) El algoritmo de Aceptación y Rechazo consiste en dos pasos:

- 1) Generar un candidato  $v_i^c$  de la distribución generadora de candidatos,  $p_G(v_i|\alpha_r^*, \gamma_r^*)$ .
- 2) Si  $u \leq \frac{p(v_i^c|resto)}{Mp_G(v_i^c; \alpha_r^*, \gamma_r^*)}$  aceptar  $v_i^c$  como muestra de  $p(v_i|resto)$ . Si el candidato  $v_i^c$  es rechazado, regresar al paso 1).

En la Figura 3.1 se presenta un ejemplo ilustrativo del funcionamiento del Algoritmo de Aceptación y Rechazo para el muestreo de  $V_i|resto$  para  $A = 0.5251$ ,  $B = 0.0121$ ,

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

$k = 3.31$ ,  $\alpha_r^* = 4.31$ ,  $\gamma_r^* = 2.3676$  y  $v^* = 5.03$ . Utilizando el Algoritmo de Aceptación y Rechazo dado en 1 y las condiciones anteriormente descritas en los pasos *a-d*, se genera un candidato  $v_i^c$  de la distribución generadora de candidatos, y se decide si es aceptado o rechazado, se realizan 1000 iteraciones del proceso con la finalidad de mostrar como funciona el algoritmo, obteniéndose los resultados mostrados en la Figura 3.1, donde se identifican con  $\circ$  los candidatos que son aceptados y con  $\times$  los candidatos rechazados.



**Figura 3.1:** Se presenta las densidades  $p(v|resto)$ ,  $g(v)$  así como  $Mg(v)$ . Además 1000 realizaciones de  $(G, V)$ , con  $G = R^2$ ,  $R \sim \Gamma(\alpha_r^*, \beta_r^*)$  y  $V = U \times M \times g(v)$ . Si  $V < p(v|resto)$  se acepta el candidato y por lo tanto  $v^c = v$ .

#### 5. $p(\omega^2|resto)$

Por otra parte, en el caso de  $\omega^2$  se tiene también que no es posible identificar el núcleo de alguna distribución univariada conocida:

$$p(\omega^2|resto) \propto (\omega^2)^{-\left(\frac{n}{2}+1\right)} \exp \left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left( \frac{\hat{e}_i}{\omega} \sqrt{\frac{v_i}{k}} - \rho z_i \right)^2 \right\}.$$

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

Se propone utilizar el algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de  $\omega^2$ . Dado que  $\omega^2 > 0$ , sea  $\vartheta^{(1)} = \log(\omega^2)$ ,  $\vartheta^{(1)}$  tiene soporte en  $\mathbb{R}$ . La función de densidad de  $\vartheta^{(1)}$  puede ser obtenida usando el método de transformación ([Casella y Berger, 2002](#)) y está dada por:

$$p(\vartheta^{(1)}|resto) \propto p(\omega^2|resto) \exp\{\vartheta^{(1)}\}.$$

En el algoritmo Metropolis con Caminata Aleatoria se genera  $\vartheta^{(1)}$  eligiendo un núcleo de transición propuesto para añadir ruido al estado actual. Suponiendo que el valor actual de  $\vartheta^{(1)}$  es  $\vartheta_{old}^{(1)}$  se desea actualizar su valor en la siguiente iteración a  $\vartheta_{new}^{(1)}$ , entonces se hace lo siguiente:

- a) Muestrear  $\vartheta^{(1)}$ ,  $\vartheta^{(1)} = \vartheta_{old}^{(1)} + \psi_1$ , donde  $\psi_1 \sim N(0, \eta_1^2)$ .
- b) Muestrear  $u$ ,  $U \sim U(0, 1)$ .
- c) Si  $u < \frac{p(\vartheta^{(1)}|resto)}{p(\vartheta_{old}^{(1)}|resto)}$  entonces  $\vartheta_{new}^{(1)} = \vartheta^{(1)}$ , de otro modo  $\vartheta_{new}^{(1)} = \vartheta_{old}^{(1)}$ .

Una vez que se ha obtenido  $\vartheta_{new}^{(1)}$ , calcular  $\omega^{2,new} = \exp\{\vartheta_{new}^{(1)}\}$ . El parámetro  $\eta_1^2$  se puede modificar para obtener una tasa de aceptación óptima.

#### 6. $p(\rho|resto)$

De igual forma, para  $\rho$ , vea que la distribución condicional no tiene forma cerrada

$$p(\rho|resto) \propto \left( \frac{1}{\sqrt{1-\rho^2}} \right)^n \exp \left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left( \frac{\hat{e}_i}{\omega} \sqrt{\frac{v_i}{k}} - \rho z_i \right)^2 \right\} \\ \times \left( \frac{1-\rho}{2} \right)^{a_0-1} \left( 1 - \frac{1-\rho}{2} \right)^{b_0-1}.$$

Se propone utilizar el algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de  $\rho$ . Se utiliza la transformación de [Fisher \(1915\)](#) definida como  $\vartheta^{(2)} = \frac{1}{2} \log \left( \frac{1+\rho}{1-\rho} \right) = \tanh^{-1}(\rho)$ , el soporte de  $\vartheta^{(2)}$  está en  $\mathbb{R}$ . Además, se consideran los hiperparámetros  $a_0 = 1$  y  $b_0 = 1$  dando paso a una a priori  $U[0, 1]$ .

### 3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

La función de densidad de  $\vartheta^{(2)}$  se puede obtener usando el método de transformación ([Casella y Berger, 2002](#)) y está dada por:

$$p(\vartheta^{(2)}|resto) \propto p(\rho|resto)\tanh(\vartheta^{(2)}).$$

En el algoritmo Metropolis con Caminata Aleatoria se generó  $\vartheta^{(2)}$  eligiendo un núcleo de transición propuesto para añadir ruido al estado actual. Suponiendo que el valor actual de  $\vartheta^{(2)}$  es  $\vartheta_{old}^{(2)}$  se desea actualizar su valor en la siguiente iteración a  $\vartheta_{new}^{(2)}$ , entonces se hace lo siguiente:

- a) Muestrear  $\vartheta^{(2)}$ ,  $\vartheta^{(2)} = \vartheta_{old}^{(2)} + \psi_2$  donde  $\psi_2 \sim N(0, \eta_2^2)$ .
- b) Muestrear  $u$ ,  $U \sim U(0, 1)$ .
- c) Si  $u < \frac{p(\vartheta^{(2)}|resto)}{p(\vartheta_{old}^{(2)}|resto)}$  entonces  $\vartheta_{new}^{(2)} = \vartheta^{(2)}$ , de otro modo  $\vartheta_{new}^{(2)} = \vartheta_{old}^{(2)}$ .

Una vez que se ha obtenido  $\vartheta_{new}^{(2)}$ , calcular  $\rho^{new} = \tanh(\vartheta_{new}^{(2)})$ . El parámetro  $\eta_2^2$  se puede modificar para obtener una tasa de aceptación óptima.

7. Finalmente, para fijar el parámetro  $k$  se sugiere utilizar una malla de puntos, por ejemplo  $k \in (2, 30)$ , y elegir el mejor modelo comparando el valor del Criterio de Información de Akaike (AIC).

Dadas las distribuciones condicionales anteriores y, como se mencionó en cada caso particular, se tiene que hacer uso de diferentes técnicas computacionales, Muestreador de Gibbs para  $\beta_0$  y  $\beta$ , se utiliza el algoritmo de Metropolis con Caminata Aleatoria para obtener muestras de  $\omega^2$  y  $\rho$ , además del algoritmo de Aceptación y Rechazo utilizado para la variable latente  $V_i$ .

Los pasos necesarios para el funcionamiento del muestreador de Gibbs, RWM y Aceptación y Rechazo para ajustar un modelo de regresión con errores aleatorios *ST* de [Azzalini y Capitanio \(2003\)](#) se describen en el Algoritmo 6. Dicho Algoritmo se implementó en el paquete estadístico *R* ([R Core Team, 2018](#)) y se presenta en el Anexo C.

3.1. Regresión Bayesiana con Errores *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*)

---

**Algoritmo 6:** Regresión Bayesiana con errores aleatorios *ST*.

```

input : El algoritmo requiere que se le proporcione  $\mathbf{y}$ ,  $\mathbf{X}$  y fijar  $k$ .
output:  $p(\boldsymbol{\theta}, \mathbf{z}, \mathbf{v} | \text{resto})$ .

1 Obtener  $\boldsymbol{\theta}^{old}$ :  $\beta_0^{old}$ ,  $\beta_j^{old}$  y  $\omega^{2,old}$  usando la función  $\mathbf{lm}()$ , fijar  $\rho^{old} = 0.5$ ,
    $\sigma_\beta^2 = 1 \times 10^6$ , y obtener  $\mathbf{z}^{old}$  y  $\mathbf{v}^{old}$  como  $n$  números aleatorios de  $N(0, 1)$  y
    $\Gamma(n, k)$  respectivamente.
2 for  $q$  in  $1 : B$  do /* con B número de iteraciones */
3   Obtener  $\hat{e}_i = y_i - (\beta_0^{old} + \mathbf{x}_i^t \boldsymbol{\beta}^{old})$ ,  $i = 1, \dots, n$ .
4   Muestrear de cada distribución posterior  $\boldsymbol{\theta} | \text{resto}$ , obtener  $\boldsymbol{\theta}^{new}$ , y actualizar
      $\hat{\mathbf{e}}$  tras actualizar cada coeficiente de regresión  $\beta_0$  y  $\beta_j$ ,  $j = 1, \dots, p$ .
5   begin
6      $\beta_0 | \beta_j^{old}, \omega^{2,old}, \rho^{old}, \mathbf{z}^{old}, \mathbf{v}^{old} \rightarrow \beta_0^{new}$ 
7      $\hat{e}_i = \hat{e}_i + (\beta_0^{old} - \beta_0^{new})$ 
8     for  $j = 1, \dots, p$  do
9        $\beta_j | \beta_0^{new}, \omega^{2,old}, \rho^{old}, \mathbf{z}^{old}, \mathbf{v}^{old} \rightarrow \beta_j^{new}$ 
10       $\hat{e}_i = \hat{e}_i + (\beta_j^{old} - \beta_j^{new})x_{ij}$ 
11    end
12     $\mathbf{z} | \beta_0^{new}, \beta_j^{new}, \omega^{2,old}, \rho^{old}, \mathbf{v}^{old} \rightarrow \mathbf{z}^{new}$ 
13    Paso de Metropolis con Caminata Aleatoria para  $\omega^2$  y  $\rho$ :
14    for  $h \in \{1, 2\}$  do
15      Muestrear de  $\vartheta^{(h)}$ ,  $\vartheta^{(h)} = \vartheta_{old}^{(h)} + \psi_h$ , con  $\psi_h \sim N(0, \eta_h^2)$ 
16      Muestrear de  $u$ ,  $U \sim U(0, 1)$ 
17      if  $u < \frac{p(\vartheta^{(h)} | \text{resto})}{p(\vartheta_{old}^{(h)} | \text{resto})}$  then
18         $\vartheta_{new}^{(h)} = \vartheta^{(h)}$ 
19      else
20         $\vartheta_{new}^{(h)} = \vartheta_{old}^{(h)}$ 
21      end
22      Para  $\omega^2$  una vez que se ha obtenido  $\vartheta_{new}^{(1)}$ , calcular
         $\omega^{2,new} = \exp \left\{ \vartheta_{new}^{(1)} \right\}$ .
23      Para  $\rho$  una vez que se ha obtenido  $\vartheta_{new}^{(2)}$ , calcular  $\rho^{new} = \tanh \left( \vartheta_{new}^{(2)} \right)$ .
24      Los parámetros  $\eta_1^2$  y  $\eta_2^2$  se pueden modificar para obtener una tasa de
        aceptación óptima.
25    end
26     $\mathbf{v} | \beta_0^{new}, \beta_j^{new}, \omega^{2,new}, \rho^{new}, \mathbf{z}^{new} \rightarrow \mathbf{v}^{new}$  /* Usando las
        especificaciones de la sección anterior. */
27  end
28 end

/* Las primeras  $m$  muestras son el calentamiento (burn-in) así que
se pueden desechar y tomar en cuenta las últimas  $B - m$  muestras
a las cuales se les aplica el adelgazamiento (thin). */

```

## 3.2. Regresión Bayesiana con Errores *ST* de [Zhu y Galbraith \(2010\)](#)

Si se considera el modelo de regresión lineal múltiple

$$y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, n; \quad \boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t, \\ e_i \sim iidST(0, \sigma^2, \alpha, k_1, k_2),$$

donde  $ST(0, \sigma^2, \alpha, k_1, k_2)$  tiene función de densidad de probabilidades dada en (2.9).

Utilizando la representación estocástica propuesta por [Naranjo et al. \(2017\)](#) y dada en (2.11), para los errores, se tiene que

$$p(e_i) = \alpha \int f(e_i | \gamma_{1i}) f(\gamma_{1i}) d\gamma_{1i} + (1 - \alpha) \int f(e_i | \gamma_{2i}) f(\gamma_{2i}) d\gamma_{2i},$$

donde

$$e_i | \gamma_{1i} \sim N(0, [2\sigma\alpha Q(k_1)]^2 \gamma_{1i}^{-1}) I(e_i \leq 0), \quad \gamma_{1i} \sim Ga(k_1/2, scale = 2/k_1), \\ e_i | \gamma_{2i} \sim N(0, [2\sigma(1 - \alpha)Q(k_2)]^2 \gamma_{2i}^{-1}) I(e_i > 0), \quad \gamma_{2i} \sim Ga(k_2/2, scale = 2/k_2).$$

Al resolver las integrales se demuestra que  $p(e_i)$  es la fdp de una distribución  $ST(0, \sigma^2, \alpha, k_1, k_2)$ .

Note que el vector de parámetros  $\boldsymbol{\theta}$  incluye los coeficientes de regresión  $\beta_0$  y  $\boldsymbol{\beta}$  y los parámetros  $\sigma^2$ ,  $\alpha$ ,  $k_1$  y  $k_2$ , es decir,  $\boldsymbol{\theta} = [\beta_0, \boldsymbol{\beta}, \sigma^2, \alpha, k_1, k_2]^t$  con  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^t$ , mientras que  $\boldsymbol{\gamma}_1 = [\gamma_{11}, \gamma_{12}, \dots, \gamma_{1n}]^t$  y  $\boldsymbol{\gamma}_2 = [\gamma_{21}, \gamma_{22}, \dots, \gamma_{2n}]^t$  son considerados pesos. Esto se conoce como “Aumentación de datos” ([Tanner y Wong, 1987](#)). Bajo esta consideración puede demostrarse que  $\mathbf{y}$  se distribuye como una *ST* de la siguiente manera:

$$y_i | \gamma_{1i} \sim N(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, [2\sigma\alpha Q(k_1)]^2 \gamma_{1i}^{-1}) I(y_i \leq \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}), \\ y_i | \gamma_{2i} \sim N(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, [2\sigma(1 - \alpha)Q(k_2)]^2 \gamma_{2i}^{-1}) I(y_i > \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}), \\ \gamma_{1i} \sim Ga(k_1/2, scale = 2/k_1), \quad \gamma_{2i} \sim Ga(k_2/2, scale = 2/k_2),$$

### 3.2. Regresión Bayesiana con Errores *ST* de Zhu y Galbraith (2010)

---

la distribución conjunta de  $y_i$ ,  $\gamma_{1i}$  y  $\gamma_{2i}$  queda definida como sigue:

$$p(y_i, \gamma_{1i}, \gamma_{2i} | \boldsymbol{\theta}) = p(y_i | \gamma_{1i}, \gamma_{2i}, \boldsymbol{\theta}) p\left(\gamma_{1i} | \frac{k_1}{2}, \frac{2}{k_1}\right) p\left(\gamma_{2i} | \frac{k_2}{2}, \frac{2}{k_2}\right),$$

entonces, la distribución de  $Y_i | \boldsymbol{\theta}$  se obtiene por integración:

$$p(y_i | \boldsymbol{\theta}) = \int p(y_i, \gamma_{1i}, \gamma_{2i} | \boldsymbol{\theta}) d\gamma_{1i} d\gamma_{2i}.$$

Resolviendo la integral anterior se puede mostrar que:

$$Y_i \sim ST(\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}, \sigma^2, \alpha, k_1, k_2), \text{ independientes.}$$

La verosimilitud aumentada se obtiene como sigue:

$$\begin{aligned} \prod_{i=1}^n p(y_i, \gamma_{1i}, \gamma_{2i} | \boldsymbol{\theta}) &= \prod_{i=1}^n \frac{\alpha}{\sqrt{2\pi[2\sigma\alpha Q(k_1)]^2 \gamma_{1i}^{-1}}} \exp\left\{-\frac{(y_i - [\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}])^2}{2[2\sigma\alpha Q(k_1)]^2 \gamma_{1i}^{-1}}\right\} \\ &\times \frac{\gamma_{1i}^{\frac{k_1}{2}-1} \exp\left\{-\frac{k_1}{2}\gamma_{1i}\right\}}{\Gamma\left(\frac{k_1}{2}\right) \left(\frac{2}{k_1}\right)^{\frac{k_1}{2}}} I[y_i \leq \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}] \\ &+ \frac{(1-\alpha)}{\sqrt{2\pi[2\sigma(1-\alpha)Q(k_2)]^2 \gamma_{2i}^{-1}}} \exp\left\{-\frac{(y_i - [\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}])^2}{2[2\sigma(1-\alpha)Q(k_2)]^2 \gamma_{2i}^{-1}}\right\} \\ &\times \frac{\gamma_{2i}^{\frac{k_2}{2}-1} \exp\left\{-\frac{k_2}{2}\gamma_{2i}\right\}}{\Gamma\left(\frac{k_2}{2}\right) \left(\frac{2}{k_2}\right)^{\frac{k_2}{2}}} I[y_i > \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}]. \end{aligned}$$

Asignando las distribuciones a priori siguientes:

$$\begin{aligned} \beta_0 | \sigma_\beta^2 &\sim N(0, \sigma_\beta^2), \\ \boldsymbol{\beta} | \sigma_\beta^2 &\sim NM(\mathbf{0}, \sigma_\beta^2 \mathbf{I}), \end{aligned}$$

### 3.2. Regresión Bayesiana con Errores $ST$ de Zhu y Galbraith (2010)

---

$$p(\sigma^2) \propto \frac{1}{\sigma^2},$$

$$\alpha|a, b \sim B(a, b).$$

Cabe señalar que para la distribución a priori de  $\beta_0$  y  $\boldsymbol{\beta}$  se fijó  $\sigma_\beta^2 = 1 \times 10^6$  dando lugar a una distribución a priori no informativa. Además, los parámetros  $k_1$  y  $k_2$  se consideran como fijos.

Con el fin de obtener estimaciones de  $\boldsymbol{\theta}$  se aplica el Teorema de Bayes utilizando la función de verosimilitud de los datos completos, la distribución a priori de  $\boldsymbol{\theta}$  y  $\boldsymbol{\Omega}$  el conjunto de hiperparámetros, el muestreador de Gibbs y un paso de Metropolis con Caminata Aleatoria.

$$p(\boldsymbol{\theta}, \gamma_1, \gamma_2 | \text{datos}) \propto p(\mathbf{y}, \gamma_1, \gamma_2 | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \boldsymbol{\Omega})$$

$$= \prod_{i=1}^n [p(y_i, \gamma_{1i}, \gamma_{2i} | \boldsymbol{\theta})] p(\beta_0 | \sigma_\beta^2) p(\boldsymbol{\beta} | \sigma_\beta^2) p(\sigma^2) p(\alpha | a, b).$$

Entonces utilizando la función de verosimilitud de los datos aumentados y las densidades a priori antes definidas, la distribución a posteriori  $p(\boldsymbol{\theta}, \gamma_1, \gamma_2 | \text{datos})$  está dada por:

$$p(\boldsymbol{\theta}, \gamma_1, \gamma_2 | \text{datos}) \propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{Q(k_1)} \frac{\gamma_{1i}^{1/2}}{\sigma} \exp \left\{ -\frac{\gamma_{1i}}{2} \left( \frac{y_i - [\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}]}{2\sigma \alpha Q(k_1)} \right)^2 \right\}$$

$$\times \frac{\gamma_{1i}^{\frac{k_1}{2}-1} \exp \left\{ -\frac{k_1}{2} \gamma_{1i} \right\}}{\Gamma\left(\frac{k_1}{2}\right) \left(\frac{2}{k_1}\right)^{\frac{k_1}{2}}} I[y_i \leq \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}]$$

$$+ \frac{1}{\sqrt{2\pi}} \frac{1}{Q(k_2)} \frac{\gamma_{2i}^{1/2}}{\sigma} \exp \left\{ -\frac{\gamma_{2i}}{2} \left( \frac{y_i - [\beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}]}{2\sigma(1-\alpha)Q(k_2)} \right)^2 \right\} \quad (3.4)$$

$$\times \frac{\gamma_{2i}^{\frac{k_2}{2}-1} \exp \left\{ -\frac{k_2}{2} \gamma_{2i} \right\}}{\Gamma\left(\frac{k_2}{2}\right) \left(\frac{2}{k_2}\right)^{\frac{k_2}{2}}} I[y_i > \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta}]$$

$$\times N(\beta_0 | 0, \sigma_\beta^2) N(\boldsymbol{\beta} | \mathbf{0}, \sigma_\beta^2 \mathbf{I}) \frac{1}{\sigma^2} B(\alpha | a, b).$$

Sea

### 3.2. Regresión Bayesiana con Errores *ST* de [Zhu y Galbraith \(2010\)](#)

---

$$\mu_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta},$$

$$\hat{e}_i = y_i - \mu_i.$$

Entonces las distribuciones condicionales completas que sirven para implementar el algoritmo del muestreador de Gibbs se pueden obtener de (3.4) y se presentan a continuación:

1.  $p(\gamma_{1i}|\text{resto})$

$$\gamma_{1i}|\text{resto} \sim Ga \left( \frac{k_1 + 1}{2}, \text{scale} = \frac{1}{\left[ \frac{\hat{e}_i^2}{2[2\sigma\alpha Q(k_1)]^2} + \frac{k_1}{2} \right]} \right), \quad \text{si } y_i \leq \mu_i.$$

2.  $p(\gamma_{2i}|\text{resto})$

$$\gamma_{2i}|\text{resto} \sim Ga \left( \frac{k_2 + 1}{2}, \text{scale} = \frac{1}{\left[ \frac{\hat{e}_i^2}{2[2\sigma(1-\alpha)Q(k_2)]^2} + \frac{k_2}{2} \right]} \right), \quad \text{si } y_i > \mu_i.$$

3.  $p(\beta_0|\text{resto})$

$$\beta_0|\text{resto} \sim N(c_{01} + c_{02}, s_{01} + s_{02}),$$

donde

$$c_{01} = \frac{\sum_{i=1}^n \gamma_{1i} y_i^* I[y_i \leq \mu_i]}{\sum_{i=1}^n \gamma_{1i} I[y_i \leq \mu_i] + \lambda_1}, \quad c_{02} = \frac{\sum_{i=1}^n \gamma_{2i} y_i^* I[y_i > \mu_i]}{\sum_{i=1}^n \gamma_{2i} I[y_i > \mu_i] + \lambda_2},$$

### 3.2. Regresión Bayesiana con Errores $ST$ de [Zhu y Galbraith \(2010\)](#)

---

$$s_{01} = \frac{[2\sigma\alpha Q(k_1)]^2}{\sum_{i=1}^n \gamma_{1i} I[y_i \leq \mu_i] + \lambda_1}, \quad s_{02} = \frac{[2\sigma(1-\alpha)Q(k_2)]^2}{\sum_{i=1}^n \gamma_{2i} I[y_i > \mu_i] + \lambda_2},$$

$$y_i^* = \hat{e}_i + \beta_0^{old}, \quad \lambda_1 = \frac{[2\sigma\alpha Q(k_1)]^2}{\sigma_\beta^2}, \quad \lambda_2 = \frac{[2\sigma(1-\alpha)Q(k_2)]^2}{\sigma_\beta^2}.$$

4.  $p(\beta_j | resto), j = 1, \dots, p$

$$\beta_j | resto \sim N(c_{j1} + c_{j2}, s_{j1} + s_{j2}),$$

donde

$$c_{j1} = \frac{\sum_{i=1}^n \gamma_{1i} y_i^* x_{ij} I[y_i \leq \mu_i]}{\sum_{i=1}^n \gamma_{1i} x_{ij}^2 I[y_i \leq \mu_i] + \lambda_1}, \quad c_{j2} = \frac{\sum_{i=1}^n \gamma_{2i} y_i^* x_{ij} I[y_i > \mu_i]}{\sum_{i=1}^n \gamma_{2i} x_{ij}^2 I[y_i > \mu_i] + \lambda_2},$$

$$s_{j1} = \frac{[2\sigma\alpha Q(k_1)]^2}{\sum_{i=1}^n \gamma_{1i} x_{ij}^2 I[y_i \leq \mu_i] + \lambda_1}, \quad s_{j2} = \frac{[2\sigma(1-\alpha)Q(k_2)]^2}{\sum_{i=1}^n \gamma_{2i} x_{ij}^2 I[y_i > \mu_i] + \lambda_2},$$

$$y_i^* = \hat{e}_i + \beta_j^{old} x_{ij}, \quad \lambda_1 = \frac{[2\sigma\alpha Q(k_1)]^2}{\sigma_\beta^2}, \quad \lambda_2 = \frac{[2\sigma(1-\alpha)Q(k_2)]^2}{\sigma_\beta^2}.$$

5.  $p(\sigma^2 | resto)$

$$\sigma^2 | resto \sim GI(a_1 + a_2 - 1, b_1 + b_2),$$

### 3.2. Regresión Bayesiana con Errores *ST* de [Zhu y Galbraith \(2010\)](#)

---

donde

$$a_1 = \frac{\sum_{i=1}^n I[y_i \leq \mu_i]}{2}, \quad a_2 = \frac{\sum_{i=1}^n I[y_i > \mu_i]}{2},$$

$$b_1 = \sum_{i=1}^n \frac{\gamma_{1i} I[y_i \leq \mu_i]}{2} \frac{\hat{e}_i^2}{[2\alpha Q(k_1)]^2}, \quad b_2 = \sum_{i=1}^n \frac{\gamma_{2i} I[y_i > \mu_i]}{2} \frac{\hat{e}_i^2}{[2(1-\alpha)Q(k_2)]^2}.$$

#### 6. $p(\alpha|resto)$

En el caso de  $\alpha$  no es posible identificar el núcleo de alguna distribución univariada conocida, pues su distribución condicional no tiene forma cerrada.

$$p(\alpha|resto) \propto \left[ \exp \left\{ -\frac{\sum_{i=1}^n \gamma_{1i} \hat{e}_i^2 I[y_i \leq \mu_i]}{2[2\sigma\alpha Q(k_1)]^2} \right\} + \exp \left\{ -\frac{\sum_{i=1}^n \gamma_{2i} \hat{e}_i^2 I[y_i > \mu_i]}{2[2\sigma(1-\alpha)Q(k_2)]^2} \right\} \right] \times \alpha^{a-1} (1-\alpha)^{b-1}.$$

Dado que  $\alpha \in (0, 1)$ , se propone utilizar la transformación  $\xi = -\log\left(\frac{1-\alpha}{\alpha}\right)$ , que es la inversa de la función Sigmoide,  $\xi$  tiene soporte en  $\mathbb{R}$ . Observe que la densidad de  $\xi$  puede ser obtenida usando el método de transformación inversa ([Casella y Berger, 2002](#)) y está dada por:

$$p(\xi|resto) \propto p(\alpha|resto) \frac{1}{1 + e^{-\xi}}.$$

En el algoritmo de Metropolis con Caminata Aleatoria se genera  $\xi$  eligiendo un núcleo de transición propuesto para añadir ruido al estado actual. Suponiendo que el valor actual de  $\xi$  es  $\xi_{old}$  se desea actualizar su valor en la siguiente iteración, a  $\xi_{new}$ , entonces se hace lo siguiente:

### 3.2. Regresión Bayesiana con Errores $ST$ de [Zhu y Galbraith \(2010\)](#)

---

- a) Muestrear  $\xi$ ,  $\xi = \xi_{old} + \psi$  donde  $\psi \sim N(0, \eta^2)$ .
- b) Muestrear  $u$ ,  $U \sim U(0, 1)$ .
- c) Si  $u < \frac{p(\xi|resto)}{p(\xi_{old}|resto)}$  entonces  $\xi_{new} = \xi$ , de otro modo  $\xi_{new} = \xi_{old}$ .

Una vez que se ha obtenido  $\xi_{new}$ , calcular  $\alpha^{new} = \frac{1}{1+e^{-\xi_{new}}}$ . El parámetro  $\eta^2$  se puede modificar para obtener una tasa de aceptación óptima.

Los pasos necesarios para la implementación del muestreador de Gibbs y RWM para ajustar un modelo de regresión con errores aleatorios  $ST$  de [Zhu y Galbraith \(2010\)](#) se describen en el Algoritmo 7.

### 3.2. Regresión Bayesiana con Errores *ST* de Zhu y Galbraith (2010)

```

Algoritmo 7: Regresión Bayesiana con errores aleatorios ST.

input : El algoritmo requiere que se le proporcione  $\mathbf{y}$ ,  $\mathbf{X}$ ,
           $\boldsymbol{\theta}^{old} = (\beta_0^{old} = 0, \beta_j^{old} = 0, \sigma^{2old} = s_y^2, \alpha = 0.5)$  y fijar  $k_1, k_2$  y
           $\sigma_\beta^2 = 1 \times 10^6$ .
output:  $p(\boldsymbol{\theta}, \gamma_1, \gamma_2 | resto)$ .

1 for  $q$  in  $1 : B$  do /* con B número de iteraciones */
2   Obtener  $\hat{e}_i = y_i - \mu_i$ , donde  $\mu_i = \beta_0^{old} + \mathbf{x}_i^t \boldsymbol{\beta}^{old}$ ,  $i = 1, \dots, n$ .
3   Muestrear de cada distribución posterior  $\boldsymbol{\theta} | resto$ , obtener  $\boldsymbol{\theta}^{new}$ , y actualizar
    $\hat{e}$  tras actualizar cada coeficiente de regresión  $\beta_0$  y  $\beta_j$ ,  $j = 1, \dots, p$ .
4   begin
5     if  $y_i \leq \mu_i$  then
6        $\gamma_{1i} | \beta_0^{old}, \beta_j^{old}, \sigma^{2old}, \alpha^{old} \rightarrow \gamma_{1i}^{new}$ 
7     else
8        $\gamma_{2i} | \beta_0^{old}, \beta_j^{old}, \sigma^{2old}, \alpha^{old} \rightarrow \gamma_{2i}^{new}$ 
9     end
10     $\beta_0 | \gamma_{1i}^{new}, \gamma_{2i}^{new}, \beta_j^{old}, \sigma^{2old}, \alpha^{old} \rightarrow \beta_0^{new}$ 
11     $\delta = \beta_0^{old} - \beta_0^{new}$ 
12     $\hat{e}_i = \hat{e}_i + \delta$ 
13    for  $j = 1, \dots, p$  do
14       $\beta_j | \gamma_{1i}^{new}, \gamma_{2i}^{new}, \beta_0^{new}, \sigma^{2old}, \alpha^{old} \rightarrow \beta_j^{new}$ 
15       $\delta = \beta_j^{old} - \beta_j^{new}$ 
16       $\hat{e}_i = \hat{e}_i + \delta x_{ij}$ 
17    end
18     $\sigma^2 | \gamma_{1i}^{new}, \gamma_{2i}^{new}, \beta_0^{new}, \beta_j^{new}, \alpha^{old} \rightarrow \sigma^{2new}$ 
19    Paso de Metropolis con Caminata Aleatoria:
20    begin
21      Muestrear de  $\xi$ ,  $\xi = \xi_{old} + \psi$ , con  $\psi \sim N(0, \eta^2)$ 
22      Muestrear de  $u$ ,  $U \sim U(0, 1)$ 
23      if  $u < \frac{p(\xi | resto)}{p(\xi_{old} | resto)}$  then
24         $\xi_{new} = \xi$ 
25      else
26         $\xi_{new} = \xi_{old}$ 
27      end
28      Una vez que se ha obtenido  $\xi_{new}$ , calcular  $\alpha^{new} = \frac{1}{1 + e^{-\xi_{new}}}$ . El
      parámetro  $\eta^2$  se puede modificar para obtener una tasa de aceptación
      óptima.
29    end
30  end
31 end

/* Las primeras  $m$  muestras son el calentamiento (burn-in) así que
se pueden desechar y tomar en cuenta las últimas  $B - m$  muestras
a las cuales se les aplica el adelgazamiento (thin). */

```

## 3.3. Estudios de Simulación

En esta sección se presentan los resultados de un pequeño estudio de simulación. El objetivo es verificar, utilizando la simulación, si la metodología propuesta funciona de forma satisfactoria al ajustar diferentes modelos para un conjunto de datos generados de forma sintética.

Con la finalidad de tener un conjunto de datos simulados en la misma escala que el conjunto de datos reales utilizado en la aplicación presentada en el Capítulo 4 para la metodología propuesta, se utilizarán valores cercanos de los parámetros de regresión y los parámetros del modelo obtenidos por el ajuste de [Azzalini y Capitanio \(2003\)](#) para el conjunto de datos de la compañía Martin Marietta para simular una nueva variable respuesta  $Y^*$  como sigue:

$$y_i^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i, \quad i = 1, 2, \dots, n$$
$$e_i \sim iidST(0, \omega^2, \lambda, k)$$

donde los parámetros verdaderos serán:

$$\begin{pmatrix} \beta_0 = 0.05 \\ \beta_1 = 1.70 \\ \beta_2 = -2.40 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} \omega^2 = 0.04 \\ \rho = 0.85 \\ k = 3.50 \end{pmatrix}$$

La covariable  $x_1$  se toma del conjunto de datos Martin Marietta, la cual representa el índice CRSP que proporciona la tasa de rendimiento mensual en la cartera del mercado de Nueva York para el período que va de 1982 a 1986. Dicha covariable consta de 60 observaciones por lo que  $n = 60$  para la simulación. La covariable  $x_2$  proviene de la simulación de  $n$  números aleatorios de una distribución  $U(0, 1)$ .

Finalmente los  $e_i$ ,  $i = 1, 2, \dots, 60$ , fueron simulados utilizando la función `rst()` la cual fue programada en el paquete estadístico *R* ([R Core Team, 2018](#)) y se presenta como parte del Anexo C. Con lo anterior es posible simular una nueva variable respuesta  $y_i^*$ ,  $i = 1, 2, \dots, 60$ .

El Algoritmo 8 resume los pasos necesarios para llevar a cabo el estudio de simulación descrito anteriormente. Se ajustaron, además, los modelos de RLM *Normal*, *t-Student*, *BSN* y *ST* de la librería **SN** ([Azzalini, 2018](#)), para comparar el ajuste de los modelos a los datos de interés.

En la Tabla 3.1 se reportan los resultados para los estimadores puntuales (obtenidos como la media posterior de las 100 réplicas Monte Carlo) de los parámetros de regresión

### 3.3. Estudios de Simulación

$\beta_0$ ,  $\beta_1$  y  $\beta_2$  así como su desviación estándar entre las réplicas. Se presentan también la correlación entre los valores observados ( $y^*$ ) y los valores predichos ( $\hat{y}^*$ ), el CME y al AIC como una medida de bondad de ajuste de los modelos, dichos estadísticos se obtuvieron como el promedio de las 100 réplicas Monte Carlo. Claramente se observa una capacidad predictiva del modelo propuesto *ST* superior al resto y una bondad de ajuste equiparable a los resultados que muestra el modelo *ST* de la librería **SN**. La Figura 3.2 presenta el histograma de residuales y el gráfico de dispersión de los valores observados vs predichos, obtenidos con los estimadores puntuales de la Tabla 3.1, mientras que en la Figura 3.3 se presentan las densidades posteriores de los parámetros estimados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\omega}$  y  $\hat{\rho}$  del modelo de regresión con errores aleatorios *ST* para el conjunto de datos simulados.

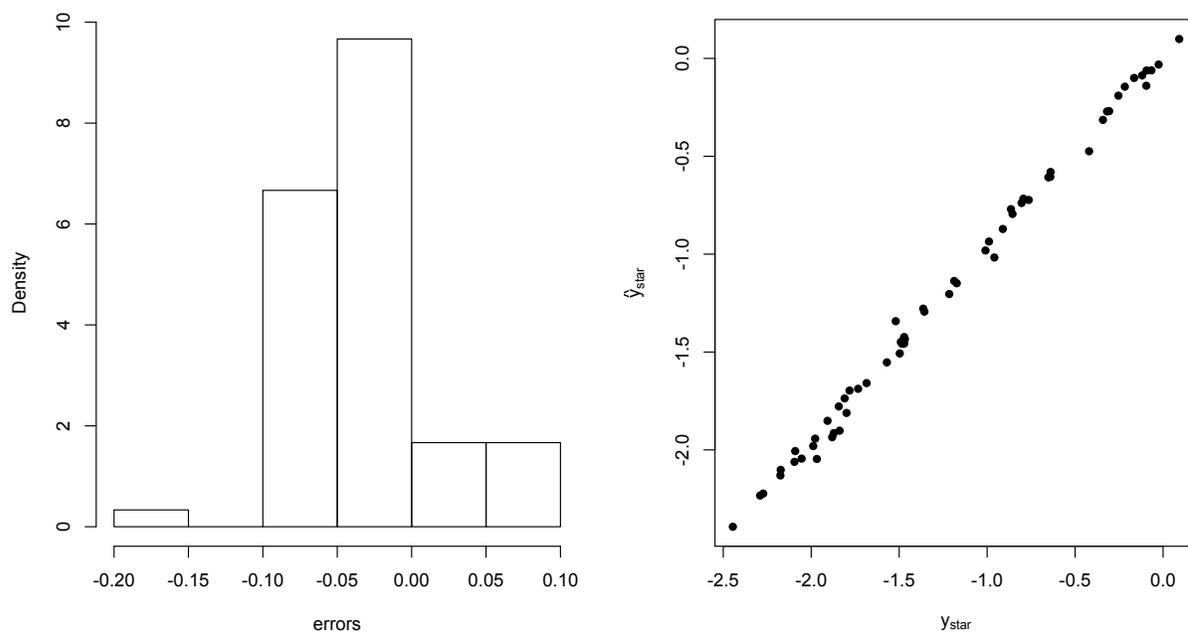
**Algoritmo 8:** Algoritmo para la realización de las Simulaciones.

- 1 Elegir el tamaño de muestra  $n$ .
- 2 Construir  $\mathbf{X}$ .
- 3 Fijar los parámetros verdaderos  $\boldsymbol{\beta}$ ,  $\omega^2$ ,  $\lambda$  y  $k$ .
- 4 Simular los  $e_i \sim iidST(0, \omega^2, \lambda, k)$ ,  $i = 1, 2, \dots, n$ .
- 5 Simular la variable respuesta  $Y^*$  como  $y_i^* = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i$ .
- 6 Ajustar el modelo de regresión con errores aleatorios *Normal* usando la función **lm()** y obtener la correlación entre los valores simulados de la respuesta y los valores ajustados,  $Corr(y^*, \hat{y}^*)$ , el Cuadrado Medio del Error (CME) y el Criterio de Información de Akaike (AIC).
- 7 **for**  $j$  *in*  $1 : M$  **do** /\* con  $M$  número de réplicas Monte Carlo \*/
- 8     Ajustar el modelo de regresión con errores aleatorios *ST* usando el Algoritmo 6 y obtener estimadores puntuales para  $\beta_0$ ,  $\boldsymbol{\beta}$  y el resto de parámetros del modelo.
- 9     Ajustar el modelo de regresión con errores aleatorios *t-Student* usando el Algoritmo 4 y obtener estimadores puntuales para  $\beta_0$ ,  $\boldsymbol{\beta}$  y el resto de parámetros del modelo.
- 10     Ajustar el modelo de regresión con errores aleatorios *SN* usando el Algoritmo 5 y obtener estimadores puntuales para  $\beta_0$ ,  $\boldsymbol{\beta}$  y el resto de parámetros del modelo.
- 11     Calcular la correlación entre los valores simulados de la respuesta y los valores ajustados,  $Corr(y^*, \hat{y}^*)$ , para cada uno de los modelos de regresión ajustados.
- 12     Calcular el CME para cada uno de los modelos de regresión ajustados.
- 13     Calcular el AIC para cada uno de los modelos de regresión ajustados.
- 14 **end**
- 15 Para obtener los estimadores de los parámetros, los modelos se ajustaron usando los datos completos y las inferencias se hicieron con base en 5,000 muestras (obtenidas después de descartar 5,000 muestras como calentamiento) de la densidad posterior y  $M = 100$  réplicas Monte Carlo. Los valores estimados se obtuvieron con respecto a la media de las 100 réplicas y también se obtuvo la desviación estándar de los estimadores entre las réplicas.

### 3.3. Estudios de Simulación

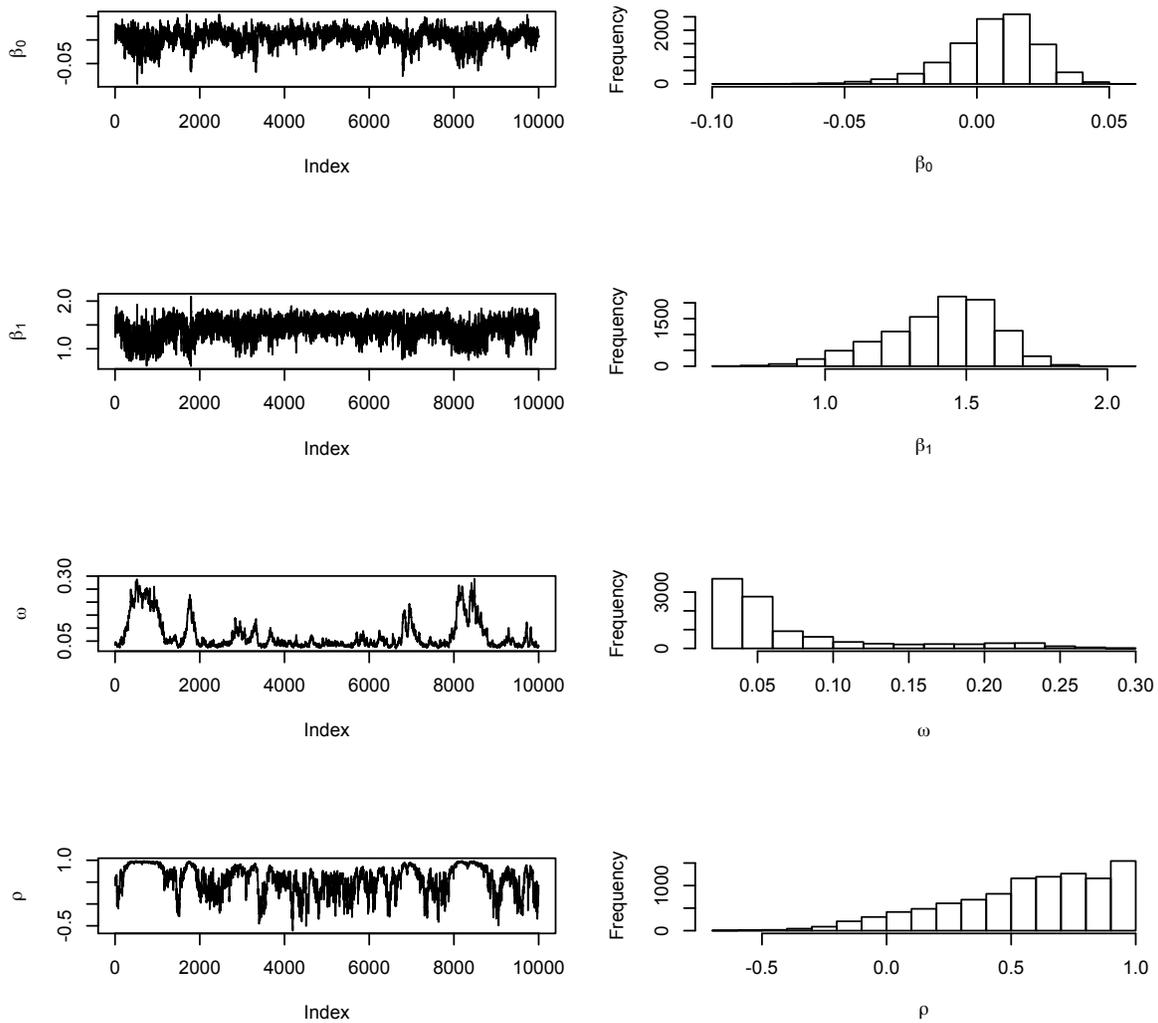
**Tabla 3.1:** Comparación del ajuste de los modelos de regresión *BST*, *Normal*, *t-Student*, *BSN* y *ST* de la librería **SN** para un conjunto de datos simulados.

$\hat{\theta}_i$ / Modelo	Normal	t-Student	BSN	ST librería SN	ST
$\hat{\beta}_0$	0.0235 (0.0118)	0.0214 (0.0019)	0.0238 (0.0006)	0.0294 (0.0117)	0.0588 (0.0047)
$\hat{\beta}_1$	1.3977 (0.1397)	1.5394 (0.0209)	1.4034 (0.0060)	1.5773 (0.0954)	1.4595 (0.0630)
$\hat{\beta}_2$	-2.3679 (0.0194)	-2.3770 (0.0032)	-2.3684 (0.0011)	-2.3772 (0.0136)	-2.3724 (0.0018)
$\hat{\omega}$	—	—	—	0.0328	0.0545
$\hat{\rho}$	—	—	0.5194	1.1415	0.7689
$\hat{k}$	—	—	—	2.4977	3.5000
$Corr(y, \hat{y})$	0.9981	0.9981	0.9981	0.9980	0.9981
$CME$	0.0020	0.0020	0.0019	0.0028	0.0126
$AIC$	-197.1043	-201.3885	-195.5449	-204.8292	-202.3118



**Figura 3.2:** Histograma de residuales (izquierda) y gráfico de dispersión de los valores observados vs valores predichos del modelo de regresión con errores aleatorios *ST* para el conjunto de datos simulados.

### 3.3. Estudios de Simulación



**Figura 3.3:** Gráfico de las densidades posteriores de los parámetros estimados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\omega}$  y  $\hat{\rho}$  del modelo de regresión con errores aleatorios *ST* para el conjunto de datos simulados. Se presenta, la Traza (lado izquierdo) y un Histograma (lado derecho) de la muestra MCMC generada para cada uno de los parámetros del modelo.

# Capítulo 4

## Aplicación con datos reales

Una aplicación del modelo propuesto consistió en ajustar el modelo de regresión  $ST$  a un conjunto de datos en el área de finanzas, el cual se describe a continuación.

### 4.1. Conjunto de Datos de Rendimientos Financieros de la compañía Martin Marietta

Los datos provienen de dos compañías cuyas acciones ordinarias cotizan en la Bolsa de Nueva York y son reportados por el Centro de Investigación en Precios de Seguridad (CRSP). Dichos datos fueron utilizados por primera vez por [Butler \*et al.\* \(1990\)](#) en un estudio sobre “Estimación robusta y parcialmente adaptativa de los modelos de regresión” en el que muestran los posibles peligros de aplicar MCO a datos con distribución no Gaussiana para obtener estimaciones de parámetros de regresión y hacer un examen de la influencia de los valores atípicos.

En dicho artículo se propone un enfoque de estimación parcialmente adaptativa para la estimación de modelos de regresión bajo diversos supuestos de distribución y presenta un ejemplo ilustrativo con un conjunto de datos leptokurtóticos de la compañía Martin Marietta obteniendo diferencias significativas en las estimaciones por MCO y los estimadores parcialmente adaptativos.

Un subconjunto de estos datos es utilizado posteriormente en un trabajo de [Azzalini y Capitanio \(2003\)](#) en el que estudian un procedimiento general para perturbar una densidad multivariada a través de la simetría y así generar un conjunto completo de densidades no simétricas, poniendo énfasis en la distribución  $t$  *Asimétrica* multivariada.

En dicho trabajo se presenta un problema de regresión lineal simple, del tipo  $y = \beta_0 +$

## 4.1. Conjunto de Datos de Rendimientos Financieros de la compañía Martin Marietta

---

$x\beta_1 + \varepsilon$ , donde  $\varepsilon \sim ST(0, \omega^2, \alpha, k)$ . Aquí, la variable de respuesta ( $y$ ) representa los rendimientos en exceso de la compañía Martin Marietta los cuales están relacionados con el índice CRSP ( $x$ ) de la tasa de rendimiento en exceso en el mercado de valores de Nueva York en general.

Finalmente, en [Arellano-Valle y Azzalini \(2013\)](#) abordan principalmente la cuestión de la elección de la parametrización de la distribución *t Asimétrica* y, más generalmente, de la selección de las cantidades de interés asociadas a esta distribución y presentan un ejemplo numérico, con el mismo conjunto de datos, sobre la base de los argumentos presentados por los anteriores autores.

El conjunto de datos utilizado en este artículo para mostrar una aplicación del modelo de regresión con errores aleatorios *ST* propuesto es incluido en la Tabla 4.1 y se encuentra disponible en la librería **hett** ([Taylor, 2018](#)) del paquete estadístico *R* ([R Core Team, 2018](#)).

Para el ajuste de los modelos de regresión se consideran las ecuaciones estimadas que asumen la relación habitual entre el rendimiento en exceso para un valor determinado y el rendimiento en exceso en la cartera del mercado, esto es,

$$(R_i - R_f) = \beta_0 + \beta_1 (R_m - R_f) + e_i,$$

donde:

$R_i$ : Tasa de rendimiento sobre el patrimonio de la empresa  $i$ ,

$R_m$ : Tasa de rendimiento de la cartera de mercado,

$R_f$ : Tasa libre de riesgo,

$e_i$ : Término de error aleatorio.

Para la estimación de los coeficientes de regresión del modelo de mercado se utilizaron 5 años de rendimientos mensuales ( $R_i$ ) de la compañía Martin Marietta, según lo reportado por el CRSP. El rendimiento libre de riesgo ( $R_f$ ) es la tasa del Tesoro a 30 días. Finalmente, el índice ponderado de valor (CRSP) con dividendos proporciona el rendimiento en la cartera del mercado ( $R_m$ ).

De esta manera se tienen las variables  $y$  y  $x$  presentadas en la Tabla 4.1 en las columnas Martin Marietta y Value Wighted CRSP Index respectivamente. Posteriormente se presentan los resultados obtenidos del ajuste de los modelos de regresión desarrollados en este trabajo con la finalidad de obtener estimaciones sobre la variación relativa de la rentabilidad del activo respecto al mercado en que cotiza.

**Tabla 4.1:** Datos: Rendimientos en Exceso, Compañía Martin Marietta (*Butler et al., 1990*).

Date	Martin Marietta <sup>(1)</sup>	Value Weighted CRSP Index <sup>(2)</sup>
Jan.1982	-0.1365	-0.0300
Feb.1982	-0.0769	-0.0584
Mar.1982	-0.0575	-0.0181
Apr.1982	0.0526	0.0306
May1982	-0.0449	-0.0397
June1982	-0.0859	-0.0295
Jul.1982	-0.0742	-0.0316
Aug.1982	0.6879	0.1176
Sept.1982	-0.0770	0.0075
Oct.1982	0.0850	0.1098
Nov.1982	0.0030	0.0408
Dec.1982	0.0754	0.0095
Jan.1983	-0.0412	0.0301
Feb.1983	-0.0890	0.0221
Mar.1983	0.2319	0.0269
Apr.1983	0.1087	0.0655
May1983	0.0375	-0.0030
June1983	0.0958	0.0325
July1983	0.0174	-0.0374
Aug.1983	-0.0724	0.0049
Sept.1983	0.0750	0.0105
Oct.1983	-0.0588	-0.0257
Nov.1983	-0.0620	0.0186
Dec.1983	-0.0378	-0.0155
Jan.1984	0.0169	-0.0165
Feb.1284	-0.0799	-0.0440
Mar.1984	-0.0147	0.0094
Apr.1984	0.0106	-0.0028
May1984	-0.0421	-0.0591
June1984	-0.0036	0.0158
July1984	0.0876	-0.0238
Aug.1984	0.1025	0.1031
Sept.1984	-0.0499	-0.0065
Oct.1984	0.1953	-0.0067
Nov.1984	-0.0714	-0.0167
Dec.1984	0.0469	0.0188
Jan.1985	0.1311	0.0733
Feb.1985	0.0461	0.0105
Mar.1985	-0.0328	-0.0070

Continua en la siguiente página.

#### 4.1. Conjunto de Datos de Rendimientos Financieros de la compañía Martin Marietta

Tabla 4.1 – Continuación de la página anterior.

Date	Martin Marietta	Value Weighted CRSP Index
Apr.1985	-0.0096	-0.0099
May1985	0.1272	0.0521
June1985	-0.0077	0.0117
July1985	0.0165	-0.0099
Aug.1985	-0.0150	-0.0102
Sept.1985	-0.1479	-0.0428
Oct.1985	-0.0065	0.0376
Nov.1985	0.0390	0.0628
Dec.1985	0.0223	0.0391
Jan.1986	-0.0690	0.0002
Feb.1986	0.1338	0.0688
Mar.1986	0.1458	0.0486
Apr.1986	0.0063	-0.0174
May1986	0.0692	0.0460
June1986	-0.0239	0.0100
July1986	-0.0568	-0.0594
Aug.1986	0.0814	0.0680
Sept.1986	-0.0889	-0.0839
Oct.1986	-0.0887	0.0481
Nov.1986	0.1037	0.0136
Dec.1986	-0.1163	-0.0322

(1) Tasa de rendimiento en exceso de la compañía.

(2) Tasa de rendimiento en exceso del mercado de valores de Nueva York.

La Tabla 4.2 presenta las principales estadísticas descriptivas del conjunto de datos Martin Marietta.

**Tabla 4.2:** Estadísticas descriptivas. Conjunto de datos Martin Marietta.

	Martin Marietta excess returns	Value weighted CRSP index
Min.	-0.14790	-0.083900
1st Qu.	-0.05960	-0.017575
Median	-0.00505	0.008450
Mean	0.01689	0.008778
3rd Qu.	0.07510	0.033775
Max.	0.68790	0.117600
Sd.	0.1208315	0.04239672

## 4.2. Ajuste del modelo *Normal*

---

### 4.2. Ajuste del modelo *Normal*

Considere el modelo de regresión lineal simple

$$y_i = \beta_0 + \beta_1 x + e_i, \quad i = 1, 2, \dots, n,$$
$$e_i \sim iidN(0, \sigma^2),$$

donde la respuesta ( $y$ ) representa la tasa de rendimiento en exceso mensual de la compañía Martin Marietta y la covariable ( $x$ ) es el índice CRSP que proporciona la tasa de rendimiento mensual en la cartera del mercado de Nueva York, ambas tasas comprenden el período que va de 1982 a 1986.

Se realizó el ajuste del modelo anterior obteniéndose los resultados mostrados en la Tabla 4.3 para la estimación de los coeficientes de regresión.

**Tabla 4.3:** Modelo clásico de regresión lineal.

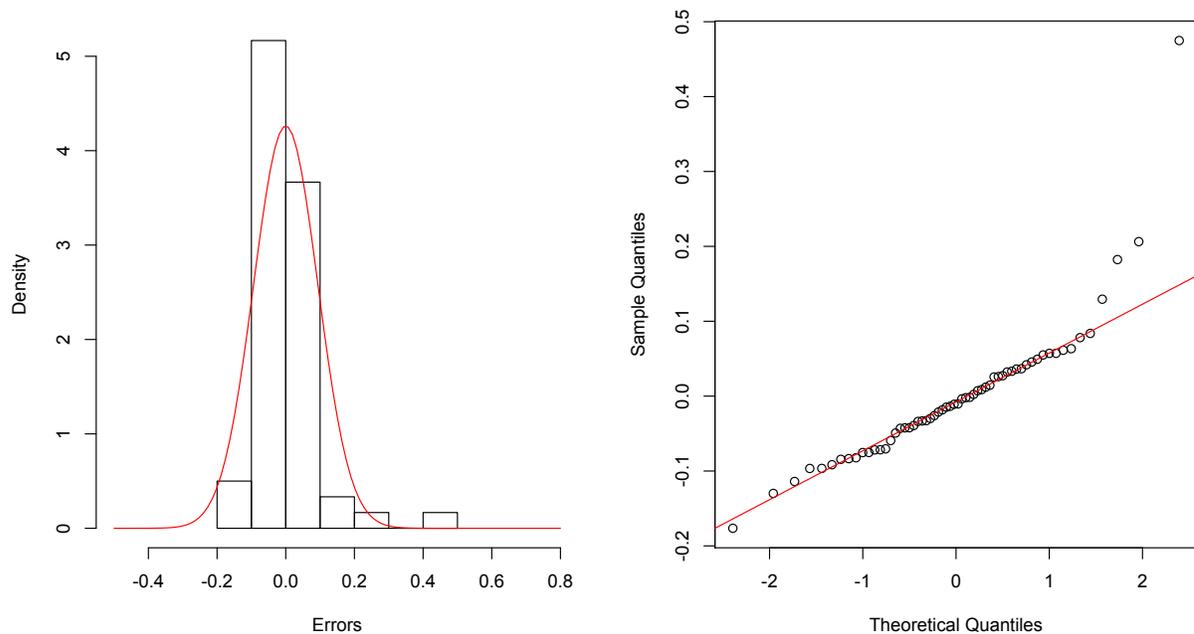
	<i>Dependent variable:</i>
	y
X	1.8025*** (0.2899)
Constant	0.0011 (0.0125)
Observations	60
R <sup>2</sup>	0.400
Adjusted R <sup>2</sup>	0.390
Residual Std. Error	0.094 (df = 58)
F Statistic	38.669*** (df = 1; 58)
AIC	-108.9898

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

La Figura 4.1 presenta un histograma de los residuales obtenidos con el ajuste del modelo *Normal*, fue superpuesta la curva de una distribución *Normal* con la misma media y desviación estándar de los residuales. También se observa la gráfica Q-Q que compara los cuantiles de la distribución observada para los residuales con los cuantiles teóricos de una distribución *Normal* con la misma media y desviación estándar.

Del histograma de la Figura 4.1 puede observarse que la curva *Normal* no ajusta a los

## 4.2. Ajuste del modelo *Normal*



**Figura 4.1:** Se muestra histograma y curva normal (izquierda) y gráfica Q-Q (derecha) de los errores del MCRL.

errores, principalmente en la cola derecha, donde claramente se distinguen valores que no son capturados por la distribución *Normal*, es decir, puede asumirse que las colas de la distribución *Normal* parecen ser muy livianas para capturar toda la información disponible del conjunto de datos de interés.

De la gráfica Q-Q de la Figura 4.1 puede observarse que los errores no se ajustan completamente a la recta de los cuantiles teóricos, principalmente en la cola derecha de la distribución observada, donde claramente se distinguen valores extremos, es decir, algunos puntos muy alejados de la recta.

Con el análisis gráfico anterior puede decirse que los errores obtenidos en el ajuste del MCRL no siguen una distribución *Normal* y por tanto el conjunto de datos Martin Marietta no se ajusta a una distribución *Normal* debido a las razones expuestas tanto en el histograma como en la gráfica Q-Q de los errores. Con la finalidad de tener más evidencias que sustenten la veracidad del planteamiento anterior, en la Tabla 4.4 se presentan los resultados de algunos métodos formales (analíticos) para probar normalidad en el conjunto de datos de interés. Las pruebas de normalidad presentadas han sido ampliamente utilizadas y han mostrado resultados confiables.

Para todas las pruebas utilizadas se tiene el siguiente juego de hipótesis:

$$H_0 : e \sim N(0, \sigma^2).$$

$$H_1 : e \not\sim N(0, \sigma^2).$$

### 4.3. Ajuste del modelo *BST*

---

Se usó un nivel de significancia de  $\alpha = 0.05$  y el criterio de decisión es:

Si el  $p - valor \leq \alpha$  Se rechaza  $H_0$ .

Si el  $p - valor > \alpha$  No se rechaza  $H_0$ .

**Tabla 4.4:** Pruebas formales de Normalidad para los errores del MCRL.

Prueba	Estadístico de Prueba	p-valor	Decisión
Shapiro-Wilk	$W = 0.82461$	5.864e-07	Se rechaza $H_0$
Lilliefors (ks.test corregido)	$D = 0.14921$	0.001973	Se rechaza $H_0$
Anderson-Darling	$A = 1.8602$	8.277e-05	Se rechaza $H_0$
Cramer-von Mises	$W = 0.27405$	0.0006291	Se rechaza $H_0$
Jarque-Bera	$JB = 284.55$	2.2e-16	Se rechaza $H_0$

Como puede observarse en la Tabla 4.4, en todas las pruebas de Normalidad utilizadas se tienen valores de los estadísticos para los cuales el  $p - valor$  determina rechazar la  $H_0$  y por lo tanto existe evidencia estadística que permite concluir que los residuales obtenidos con el ajuste del MCRL no siguen una distribución *Normal* y por tanto el conjunto de datos Martin Marietta tampoco se ajusta a una distribución *Normal*.

### 4.3. Ajuste del modelo *BST*

Considere el modelo de regresión lineal simple

$$y_i = \beta_0 + \beta_1 x + e_i, \quad i = 1, 2, \dots, n,$$
$$e_i \sim iidST(0, \omega^2, \lambda, k),$$

donde  $ST(0, \omega^2, \lambda, k)$  tiene la forma de (2.8). La respuesta ( $y$ ) representa la tasa de rendimiento en exceso de la compañía Martin Marietta y la covariable ( $x$ ) es el índice CRSP que proporciona la tasa de rendimiento en la cartera del mercado de Nueva York.

Se realizó el ajuste del modelo anterior usando la función `lm.st()` que se implementó en el paquete estadístico *R* (R Core Team, 2018) y se muestra en el Anexo C, obteniéndose los resultados de la Tabla 4.5 para la estimación de los coeficientes de regresión y el resto de parámetros generados por el modelo.

### 4.3. Ajuste del modelo *BST*

---

**Tabla 4.5:** Modelo de Regresión con errores *ST*.

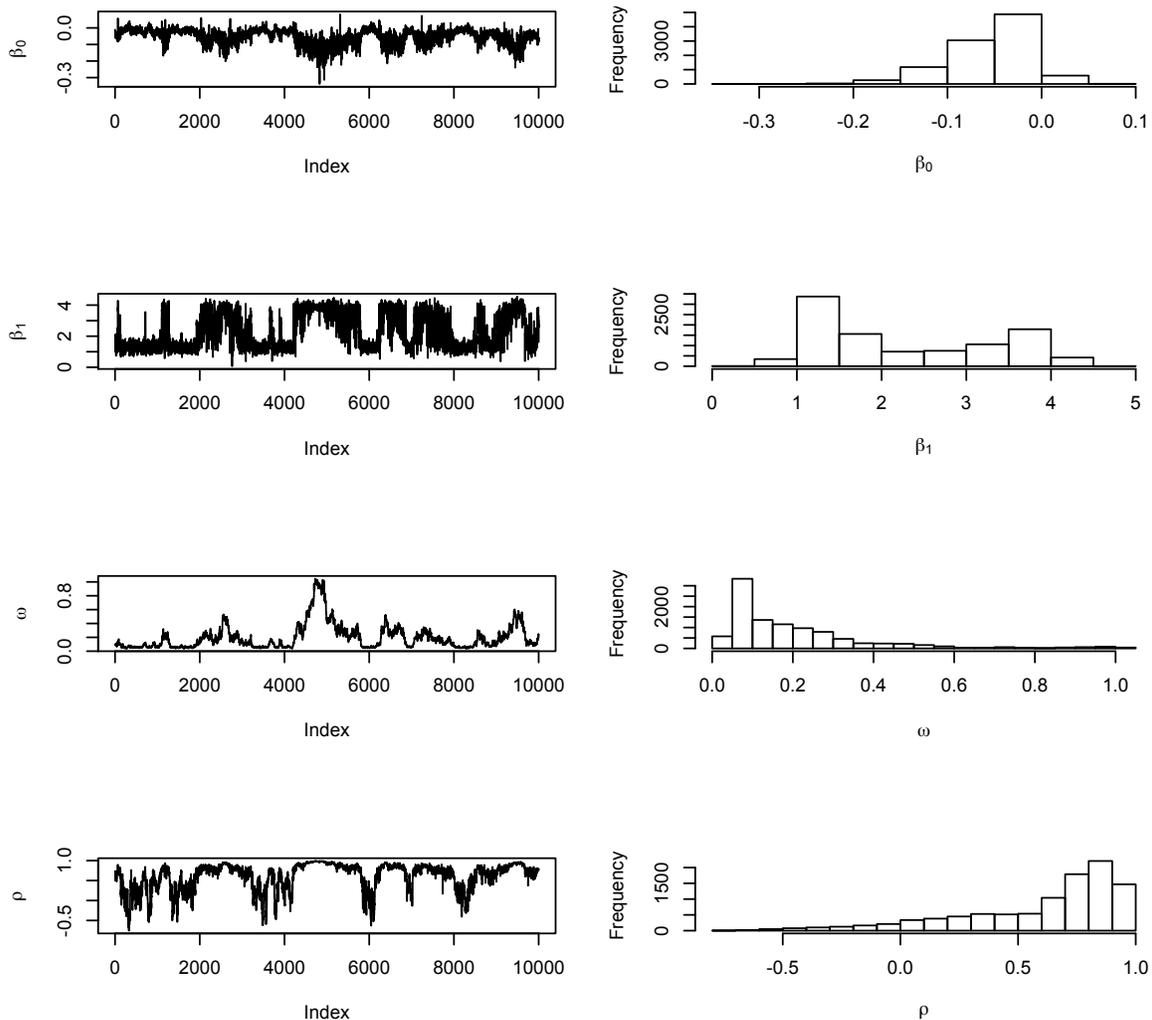
<i>Dependent variable:</i>	
y	
X	1.3030 (0.8651)
Constant	0.0280 (0.0341)
<i>Parameters generated by the model:</i>	
$\omega$	0.0618
$\rho$	0.7706
$k$	3.0000
Observations	60
<i>Corr</i> ( $y, \hat{y}$ )	0.6325
<i>CME</i>	0.0095
AIC	-133.7135

En la Figura 4.2 se presentan las densidades posteriores de los parámetros estimados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\omega}$  y  $\hat{\rho}$ , para el modelo de regresión con errores aleatorios *ST* del conjunto de datos Martin Marietta.

En la Figura 4.3 se presenta el histograma de los residuales y el gráfico de dispersión de los valores observados contra los valores predichos por el modelo *ST*.

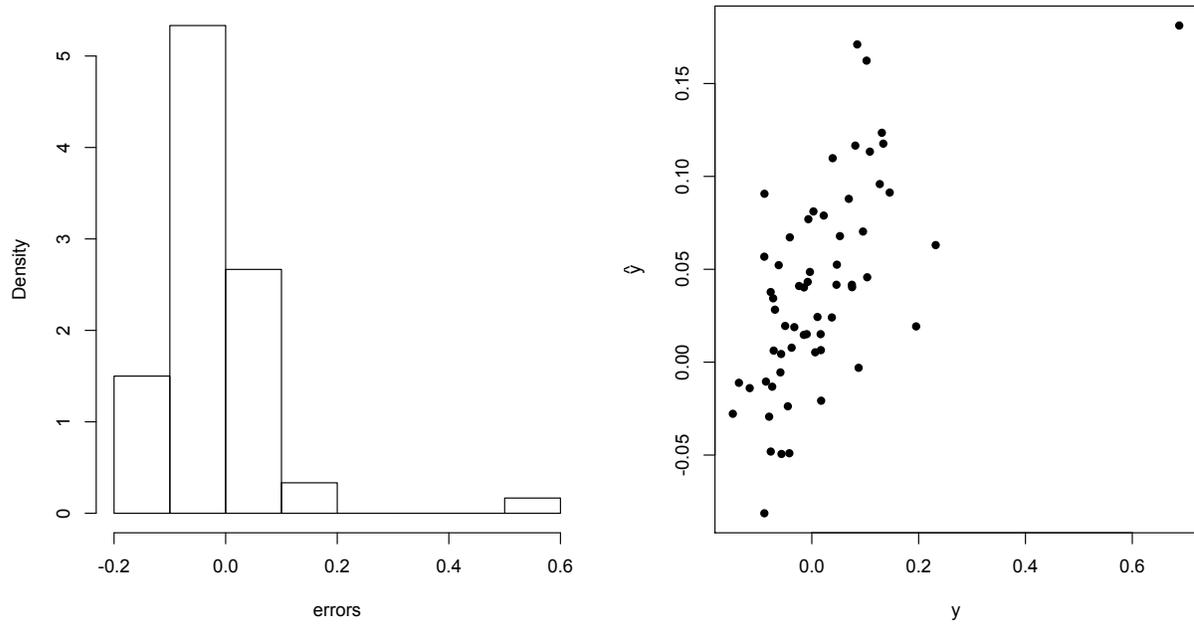
Finalmente, en la Tabla 4.6 se reportan los resultados obtenidos, para los estimadores puntuales de los parámetros de regresión  $\beta_0$  y  $\beta_1$  así como sus errores estándar, utilizando los diferentes modelos analizados en este trabajo de investigación, que incluyen, el modelo *Normal*, *t-Student*, *BSN* y el modelo propuesto *BST*, además de los resultados que ofrece el modelo *ST* de la librería **SN** (Azzalini, 2018); lo anterior con el fin de comparar el ajuste de los modelos a los datos de interés. Se presentan también la correlación entre los valores observados ( $y$ ) y los valores ajustados ( $\hat{y}$ ), el *CME* y el *AIC*, como medidas de bondad de ajuste de los modelos. Para obtener los estimadores de los parámetros y los estadísticos mencionados, las inferencias se hicieron con base en 5,000 muestras (obtenidas después de descartar 5,000 muestras como calentamiento) y adelgazamiento con cada 10 observaciones de las densidades posteriores de cada modelo.

### 4.3. Ajuste del modelo *BST*



**Figura 4.2:** Gráfico de las densidades posteriores de los parámetros estimados  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\omega}$  y  $\hat{\rho}$  del modelo de regresión con errores aleatorios *ST* para el conjunto de datos Martin Marietta. Se presenta, la Traza (lado izquierdo) y un Histograma (lado derecho) de la muestra MCMC generada para cada uno de los parámetros del modelo.

### 4.3. Ajuste del modelo *BST*



**Figura 4.3:** Histograma de residuales (izquierda) y gráfico de dispersión de los valores observados vs valores predichos por el modelo de regresión con errores aleatorios *ST* para el conjunto de datos Martin Marietta.

**Tabla 4.6:** Comparación del ajuste de los modelos de regresión *BST*, *Normal*, *t-Student*, *BSN* y *ST* de la librería *SN* para el conjunto de datos Martin Marietta.

$\hat{\theta}_i$ / Modelo	Normal	t-Student	BSN	ST librería SN	BST
$\hat{\beta}_0$	0.0011 (0.0125)	-0.0049 (0.0083)	0.0099 (0.0117)	0.0029 (0.0233)	0.0280 (0.0341)
$\hat{\beta}_1$	1.8025 (0.2899)	1.2897 (0.1993)	1.4351 (0.2818)	1.2481 (0.1901)	1.3030 (0.8651)
$\hat{\omega}$	—	—	—	0.0655	0.0618
$\hat{\rho}$	—	—	0.9657	0.8901	0.7706
$\hat{k}$	—	—	—	3.3194	3.0000
$Corr(y, \hat{y})$	0.6325	0.6325	0.6325	0.6325	0.6325
$CME$	0.0089	0.0092	0.0089	0.0124	0.0095
$AIC$	-108.9898	-135.4643	-123.9178	-136.1661	-133.7135

#### 4.4. Prueba de Bondad de Ajuste de Correlación

---

De la Tabla 4.6 observe que la  $Corr(y, \hat{y})$  es prácticamente idéntica para los cuatro modelos por lo que el gráfico de los valores observados vs los valores predichos por el modelo  $ST$  será muy similar al generado por el resto de los modelos. Sin embargo, se han presentado pruebas formales de bondad de ajuste que demuestran un problema de especificación del modelo al asumir una distribución *Normal* en los errores. De los resultados obtenidos puede asumirse que los modelos *t-Student*,  $SN$  y  $ST$  ajustan bien a los datos. En seguida se presentan dos Pruebas de Bondad de Ajuste para la distribución  $ST$ , una basada en la distribución del coeficiente de Correlación y otra basada en el cálculo del estadístico de Kolmogorov-Smirnov mediante *Bootstrap* paramétrico. Ambas pruebas demuestran que resulta correcto asumir una distribución  $ST$  en los errores para ajustar un modelo de regresión robusta con el conjunto de datos analizado.

#### 4.4. Prueba de Bondad de Ajuste de Correlación

Pérez-Rodríguez y Villaseñor (2010) desarrollaron una prueba de bondad de ajuste para la familia  $SN$  basada en las propiedades de esta familia de distribuciones y el coeficiente de correlación de la muestra, donde los valores críticos para las pruebas se obtuvieron utilizando simulación Monte Carlo, además mostraron que su prueba tiene mayor poder que las pruebas basadas en la función de distribución empírica frente a algunas distribuciones alternativas.

Maghami y Bahrami (2015) proponen una prueba de bondad de ajuste para la distribución *t Asimétrica* basandose en el estudio de Pérez-Rodríguez y Villaseñor (2010), usando la parametrización de Azzalini y Capitanio (2003). En este trabajo se utiliza la prueba antes descrita, aplicada a los errores de regresión, para probar la bondad de ajuste del modelo presentado en este capítulo para el conjunto de datos Martin Marietta.

Dada una muestra aleatoria  $y_1, \dots, y_n$  queremos probar el siguiente juego de hipótesis:

$$H_0 : f(y) = f_Y(y; \xi, \omega^2, \lambda, k) \text{ dada en (2.8).}$$

$$H_1 : f(y) \neq f_Y(y; \xi, \omega^2, \lambda, k) \text{ dada en (2.8).}$$

La prueba se basa en la propiedad 6) de la  $ST$ , si  $W \sim ST(\lambda, k)$ , entonces  $W^2 \sim F(1, k)$ , a partir de la cual se tiene lo siguiente:

Sea  $Y$  una variable aleatoria definida como

$$Y = \xi + \omega W,$$

#### 4.4. Prueba de Bondad de Ajuste de Correlación

---

donde  $Y \sim ST(\xi, \omega^2, \lambda, k)$  y  $W \sim ST(\lambda, k)$ , entonces

$$Z := (Y - \xi)^2 = \omega^2 W^2, \quad (4.1)$$

y por la propiedad 5):

$$\frac{Z}{\omega^2} := W^2 \sim F(1, k). \quad (4.2)$$

De lo anterior note que el parámetro  $\lambda$  ha sido eliminado del problema. Para un  $\xi$  y  $k$  fijos, digamos  $\xi = \xi_0$  y  $k = k_0$ ,  $Z$  tiene una distribución escalada:

$$P(Z \leq z) = G\left(\frac{z}{\omega^2}\right),$$

donde  $G$  es la función de distribución de una variable aleatoria  $F(1, k_0)$ . Dada una muestra  $y_1, \dots, y_n$  y  $\xi_0$ , calcular  $z_1, \dots, z_n$  usando (4.1). Un estimador consistente para  $P(Z \leq z)$  es la función de distribución empírica, dada por:

$$F_n(z) = \begin{cases} 0, & z < z_{(1)}, \\ i/n, & z_{(i)} \leq z < z_{(i+1)}, \\ 1, & z_{(n)} < z, \end{cases}$$

donde  $z_{(1)} \leq \dots \leq z_{(n)}$  son las estadísticas de orden de  $z'_i$ s, entonces

$$G\left(\frac{z}{\omega^2}\right) \approx F_n(z),$$

por lo tanto:

$$u = G^{-1}(F_n(z)) \approx \frac{z}{\omega^2}. \quad (4.3)$$

Una vez establecidos los  $u'_i$ s, debemos esperar una relación lineal fuerte entre los  $z'_i$ s y  $u'_i$ s bajo la  $H_0$ . Si obtenemos  $\hat{\xi}$  y  $\hat{k}$ , estimadores consistentes de  $\xi$  y  $k$ , entonces se esperaría que la relación lineal aun se mantenga. Para probar lo anterior se utiliza el coeficiente de correlación muestral, que está dado por:

$$r_n = \text{Corr}(Z, U) = \frac{\sum_{i=1}^n (Z_i - \bar{Z}_i)(U_i - \bar{U}_i)}{\sum_{i=1}^n \sqrt{(Z_i - \bar{Z}_i)^2 (U_i - \bar{U}_i)^2}}. \quad (4.4)$$

La  $H_0$  es rechazada con un nivel de significancia  $\alpha \in (0, 1)$  si  $r_n \leq C_n(\alpha)$ , donde  $C_n(\alpha)$  es tal que:

$$\alpha = \max_{\lambda} P(\text{Rechazar } H_0 | H_0) = \max_{\lambda} P(r_n \leq C_n(\alpha)).$$

#### 4.4. Prueba de Bondad de Ajuste de Correlación

La distribución de  $r_n$  bajo la hipótesis nula para cada valor fijo de  $k$  y  $\lambda$  se puede obtener mediante simulación de Monte Carlo. Tenga en cuenta que  $r_n$  es invariante a escala y que la distribución de  $Z$  en (4.1) no depende de  $\xi$ , por lo tanto, se fijan  $\xi = 0$  y  $\omega^2 = 1$ .

Note que si una muestra aleatoria proviene de una función de distribución diferente de la distribución  $ST$  para la cual la propiedad 6) no se cumple, esto significa que la ecuación (4.3) no se cumple, entonces, el coeficiente de correlación de la muestra en (4.4) no puede estar cerca de 1, por lo tanto,  $r_n$  debería ser menor que el valor crítico, ya que bajo  $H_0$  la distribución de  $r_n$  se concentrará cerca de 1.

El Algoritmo 9 presenta el procedimiento mediante el cual es posible obtener las constantes críticas para diferentes valores de  $n$  y  $k$ . Como ya se mencionó, el parámetro  $\lambda$  no influye en la prueba, Maghami y Bahrami (2015) proponen utilizar  $\lambda = 20$ .

<b>Algoritmo 9:</b> Algoritmo para obtener constantes críticas de la Prueba de Bondad de Ajuste de Correlación.	
<b>input</b>	: Fijar $n$ , $\xi = 0$ , $\omega^2 = 1$ , $\lambda = 20$ , y $k$ .
<b>output:</b>	$r_n(\alpha)$ .
1	<b>for</b> $j$ <i>in</i> $1 : B$ <b>do</b> /* con B número de muestras deseadas */
2	Simular una muestra de tamaño $n$ de $ST(\xi, \omega^2, \lambda, k)$ .
3	Obtener el EMV de $\xi$ .
4	Calcular $z_i$ , $i = 1, \dots, n$ usando (4.1).
5	Ordenar en forma ascendente las $z_i$ 's.
6	Calcular $u_i = G^{-1}(F_n(z_i))$ , $i = 1, \dots, n$ , donde $G^{-1}$ es la función de cuantiles de la distribución $F(1, k)$ .
7	Calcular el coeficiente de correlación $r_n = Corr(z_i, u_i)$ usando (4.4).
8	<b>end</b>
9	Se tienen $B$ realizaciones de $r_n$ para unos valores de $\lambda$ y $k$ dados.
10	El valor de la constante crítica $C_n(\alpha)$ , con $\alpha \in (0, 1)$ , es determinado por los cuantiles $100\alpha$ de la distribución empírica de $r_n$ obtenida en el proceso anterior.

Bajo el contexto de regresión lineal, se desea probar el siguiente juego de hipótesis:

$$H_0 : f(e) = f(e; 0, \omega^2, \lambda, k) \text{ dada en (2.8).} \quad (4.5)$$

$$H_1 : f(e) \neq f(e; 0, \omega^2, \lambda, k) \text{ dada en (2.8).}$$

Note que una vez calculados los  $z_i$ 's, el problema se reduce a probar si  $Z_1, \dots, Z_n$  es una muestra aleatoria de la distribución  $F(1, k)$ .

Por tanto, se pueden resumir los pasos necesarios para llevar a cabo una prueba de bondad de ajuste de correlación, bajo el contexto de regresión, para cualquier conjunto de datos, como se muestra en el Algoritmo 10.

#### 4.4. Prueba de Bondad de Ajuste de Correlación

**Algoritmo 10:** Algoritmo para la realización de la Prueba de Bondad de Ajuste de Correlación para los errores.

**input** :  $\mathbf{y}$ ,  $\mathbf{X}$ , conjunto de datos de tamaño  $n$ .

- 1 Ajustar el modelo  $y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i$ ,  $e_i \sim iidST(0, \omega^2, \lambda, k)$ .
- 2 Obtener los residuales  $\hat{e}_i = y_i - \hat{\beta}_0 - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}$ .
- 3 Bajo  $H_0$ ,  $f(e) = f(e; 0, \omega^2, \lambda, k)$  dada en (2.8), calcular  $z_i$ ,  $i = 1, \dots, n$  usando (4.1) y los estimadores  $\hat{\omega}^2$ ,  $\hat{\lambda}$  y  $\hat{k}$  obtenidos en el ajuste del modelo.
- 4 Ordenar en forma ascendente las  $z_i$ 's.
- 5 Calcular  $u_i = G^{-1}(F_n(z_i))$ ,  $i = 1, \dots, n$ , donde  $G^{-1}$  es la función de cuantiles de la distribución  $F(1, k)$ .
- 6 Calcular el coeficiente de correlación  $r_n = Corr(z_i, u_i)$  usando (4.4).
- 7 Para un nivel de significancia  $\alpha \in (0, 1)$ , identificar el cuantil  $C_n(\alpha)$  de la prueba estadística correspondiente para  $k$  y  $n$ , usando los pasos del Algoritmo 9.
- 8 Si  $r_n \leq C_n(\alpha)$ , rechazar  $H_0$  con un nivel de significancia  $\alpha$ .

Utilizando los coeficientes de regresión estimados de la Tabla 4.5 obtenidos en el ajuste del modelo de regresión con errores aleatorios  $ST$  para el conjunto de datos Martin Marietta, presentados en la sección anterior, se obtuvieron los residuales  $\hat{e}_i$ ,  $i = 1, \dots, n$ , con  $n = 60$ , se fijó  $\lambda = 20$  y  $k = 3.00$ . Utilizando los Algoritmos 9 y 10 descritos, se realizó la Prueba de Bondad de Ajuste de Correlación para los errores para un nivel de significancia  $\alpha = 0.05$ , obteniéndose los resultados de la Tabla 4.7.

**Tabla 4.7:** Prueba de Bondad de Ajuste de Correlación para los residuales del modelo  $BST$ .

Hipótesis	Regla de Decisión	Prueba	Decisión
$H_0 : Z \sim F(1, k)$	Rechazar $H_0$	$0.9276 > 0.8663$	No rechazar $H_0$
$H_1 : Z \not\sim F(1, k)$	si $r_n \leq C_n(\alpha)$	$r_n > C_n(\alpha)$	

Nota:  $\alpha = 0.05$

Entonces, del juego de hipótesis original en (4.5) y como consecuencia de la Tabla 4.7, se concluye que no se rechaza  $H_0$ :  $f(e) = f(e; 0, \omega^2, \lambda, k)$  dada en (2.8), por lo que el ajuste del modelo de regresión con errores  $ST$  es adecuado para el conjunto de datos Martin Marietta.

## 4.5. Prueba de Bondad de Ajuste usando *Bootstrap* Paramétrico

En esta sección se describe una técnica paramétrica que se utiliza para estimar los cuantiles de un estadístico de prueba, llamémosle  $T$ , el cual se basa en la Función de Distribución Empírica,  $F_n(z)$ , cuando la distribución hipotética es  $ST$  con valores de parámetros estimados a partir de los datos. Para las estimaciones puede utilizarse el método de Máxima Verosimilitud, sin embargo, como las expresiones analíticas no existen para estos estimadores, se deben usar métodos numéricos para calcularlos. Tenga en cuenta que cuando los parámetros desconocidos son parámetros de localidad o escala, y se estiman utilizando estimadores equivalentes (al igual que los estimadores de máxima verosimilitud), las distribuciones muestrales de las estadísticas basadas en la Función de Distribución Empírica no dependen de los valores reales de esos parámetros (Eastman y Bain, 1973). Por lo tanto, los valores de  $\xi = 0$  y  $\omega = 1$  se utilizan por simplicidad debido a que la distribución muestral de una estadística  $T$  dada es invariante a cambios en los parámetros de localidad y escala. Sin embargo, para el caso de la  $ST$ , dado que la distribución nula asintótica del estadístico de prueba depende del valor desconocido de  $\lambda$  y  $k$ , se realiza un versión de *bootstrap* paramétrico de la prueba (Maghami y Bahrami, 2015).

Pérez-Rodríguez y Villaseñor (2010) realizaron un estudio como el antes descrito para la familia  $SN$ , mientras que Maghami y Bahrami (2015) replican este procedimiento para proponer una prueba de bondad de ajuste usando *bootstrap* paramétrico para la distribución  $ST$  utilizando un conjunto de estadísticas de prueba  $T$  basadas en la  $F_n(z)$ .

Estamos interesados en probar el juego de hipótesis en (4.5) de modo que en este apartado se retoma el procedimiento descrito utilizando el estadístico de Kolmogorov-Smirnov,  $D$ , el cual se define como sigue.

Dada una muestra aleatoria  $y_i$ ,  $i = 1, \dots, n$ , calcular  $v_i = F(y_i)$ , donde  $y_1 \leq \dots \leq y_n$  y  $F(\cdot)$  es la función acumulativa de probabilidades de una  $ST$  (los parámetros desconocidos son reemplazados por sus estimadores), el estadístico de Kolmogorov-Smirnov es:

$$D_{calc} = \max(D^+, D^-), \quad (4.6)$$

donde

$$D^+ = \max_{1 \leq i \leq n} \left( \frac{i}{n} - v_i \right), \quad D^- = \max_{1 \leq i \leq n} \left( v_i - \frac{i-1}{n} \right).$$

La regla de decisión es rechazar  $H_0$  si  $D_{calc} > D_{tab}(\alpha)$ , con  $\alpha \in (0, 1)$  el nivel de significancia de la prueba.

Con la finalidad de probar (4.5) y usando el estadístico de Kolmogorov-Smirnov, el

#### 4.5. Prueba de Bondad de Ajuste usando *Bootstrap* Paramétrico

Algoritmo 11 permite obtener los valores críticos  $D_{tab}(\alpha)$  de la prueba mediante el procedimiento de *Bootstrap* paramétrico así como realizar la prueba bajo el contexto de regresión que nos ocupa en este trabajo de investigación.

<b>Algoritmo 11:</b> Algoritmo para obtener constantes críticas y realización de la Prueba de Bondad de Ajuste para los errores usando <i>Bootstrap</i> Paramétrico.	
<b>input</b> : $\mathbf{y}$ , $\mathbf{X}$ , conjunto de datos de tamaño $n$ .	
1	Ajustar el modelo $y_i = \beta_0 + \mathbf{x}_i^t \boldsymbol{\beta} + e_i$ , $e_i \sim iidST(0, \omega^2, \lambda, k)$ .
2	Obtener los residuales $\hat{e}_i = y_i - \hat{\beta}_0 - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}$ .
3	Bajo $H_0$ , $f(e) = f(e; 0, \omega^2, \lambda, k)$ dada en (2.8), calcular $D_{calc}$ utilizando la ecuación (4.6), donde $F(\cdot)$ denota la función de distribución de una $ST(0, 1, \hat{\lambda}, \hat{k})$ , $\hat{\lambda}$ y $\hat{k}$ son los estimadores obtenidos en el ajuste del modelo.
4	<b>for</b> $j$ <i>in</i> $1 : B$ <b>do</b> /* con $B = 1000$ número de muestras deseadas */
5	Simular una muestra <i>Bootstrap</i> de tamaño $n$ de $ST(0, 1, \hat{\lambda}, \hat{k})$ .
6	Dada la muestra <i>Bootstrap</i> estimar los MLE's de $\lambda$ y $k$ , llamémosle $\tilde{\lambda}$ y $\tilde{k}$ .
7	Calcular $D_{boot}$ usando (4.6), los estimadores $\tilde{\lambda}$ y $\tilde{k}$ y la muestra <i>Bootstrap</i> .
8	<b>end</b>
9	Se tienen $B = 1000$ realizaciones de $D_{boot}$ . El valor de la constante crítica $D_{tab-boot}(\alpha)$ , con $\alpha \in (0, 1)$ el nivel de significancia de la prueba, si $\alpha = 0.05$ , se obtiene como $D_{boot(950)}$ , donde $D_{boot(j)}$ , $j = 1, \dots, 1000$ denota los valores ordenados de $D_{boot}$ .
10	Se rechaza $H_0$ en (4.5), al nivel de significancia $\alpha$ si $D_{calc} > D_{boot(950)}$ .

Utilizando los coeficientes de regresión estimados de la Tabla 4.5 obtenidos en el ajuste del modelo de regresión con errores aleatorios  $ST$  para el conjunto de datos Martin Marietta, se obtuvieron los residuales  $\hat{e}_i$ ,  $i = 1, \dots, n$ , con  $n = 60$ . Utilizando el Algoritmo 11 descrito, se realizó la Prueba de Bondad de Ajuste para los errores usando *Bootstrap* paramétrico para un nivel de significancia  $\alpha = 0.05$ , obteniéndose los resultados mostrados en la Tabla 4.8.

**Tabla 4.8:** Prueba de Bondad de Ajuste para los residuales del modelo *BST* usando *Bootstrap* Paramétrico

Hipótesis	Regla de Decisión	Prueba	Decisión
$H_0 : e \sim ST(0, \omega^2, \lambda, k)$	<i>Rechazar</i> $H_0$ si	$0.8899 < 0.9934$	No rechazar $H_0$
$H_1 : e \approx ST(0, \omega^2, \lambda, k)$	$D_{calc} > D_{boot-tab}(\alpha)$	$D_{calc} < D_{boot-tab}(\alpha)$	

*Nota:*  $\alpha = 0.05$

Por lo tanto, del juego de hipótesis original en (4.5) y con los resultados de la Tabla 4.8, se concluye que no se rechaza  $H_0: e \sim ST(0, \omega^2, \lambda, k)$ , por lo que el ajuste del modelo de regresión con errores  $ST$  es adecuado para el conjunto de datos Martin Marietta.

# Capítulo 5

## Conclusiones

En la literatura reciente se ha examinado la familia de distribuciones asimétricas, la cual ofrece una gran flexibilidad para adaptarse a una extensa gama de situaciones prácticas, y aún así mantiene la capacidad de tratamiento matemático y un conjunto de propiedades formales atractivas. Algunas evidencias numéricas y la disponibilidad de software desarrollado para inferencia proporcionan soporte adicional para usar esta familia de distribuciones en casos prácticos. En este trabajo de investigación se ha mostrado que la  $ST$  es una clase de distribución asimétrica univariada que tiene características deseables por las cuales se puede considerar una generalidad a modelos que siguen otras distribuciones como la *Normal*, *t-Student*, e incluso algunas dentro de la propia familia de distribuciones asimétricas, como lo son la  $SN$ ,  $SC$ , entre otras.

Generalmente, el uso de la distribución  $ST$  en problemas de regresión lineal múltiple bajo el contexto Bayesiano, representa una herramienta estadística adecuada para tratar conjuntos de datos donde el supuesto de Normalidad en los errores no se cumple, es decir, conjuntos de datos con valores atípicos, reduciendo la influencia de dichas observaciones aberrantes en la inferencia para los parámetros de interés, debido a que esta distribución es de cola pesada y asimétrica, lo que permite tener un mayor ajuste del modelo y obtener predicciones adecuadas y de mayor utilidad que en otros caso. En el caso específico de los modelos para el comportamiento de los mercados financieros, a pesar de que el supuesto de Normalidad se ha considerado inadecuado, se ha mantenido la metodología asociada, debido principalmente, a la falta de metodologías de reemplazo adecuadas. La  $ST$  ofrece una alternativa más, que permite mantener las formulaciones del comportamiento de los mercados financieros al momento que se adopta un supuesto de distribución más realista.

Si bien, esta no siempre resultará la proposición más adecuada, pudiendo existir en algunos casos alternativas con mejores resultados, bajo las condiciones en que trabaja la  $ST$  ya mencionadas, la herramienta estadística aquí planteada contribuye a un conjunto amplio y estrechamente interconectado de resultados que están evolucionando hacia un marco bastante general dentro del cual hace falta todavía continuar la investigación.

# Referencias

- Arellano-Valle, R. B. y Azzalini, A. (2013). The centred parameterization and related quantities of the skew-t distribution. *Journal of Multivariate Analysis*, 113, 73 – 90.
- Azevedo, C. L., Bolfarine, H. y Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Computational Statistics & Data Analysis*, 55, 1, 353–365.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 171–178.
- Azzalini, A. (2018). *The R package sn: The Skew-Normal and Related Distributions such as the Skew-t (version 1.5-3)*. Università di Padova, Italia.
- Azzalini, A. y Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 3, 579–602.
- Azzalini, A. y Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 2, 367–389.
- Branco, M. D. y Dey, D. K. (2001). A General Class of Multivariate Skew-Elliptical Distributions. *Journal of Multivariate Analysis*, 79, 1, 99–113.
- Butler, R. J., McDonald, J. B., Nelson, R. D. y White, S. B. (1990). Robust and Partially Adaptive Estimation of Regression Models. *The Review of Economics and Statistics*, 72, 2, 321–327.
- Carlin, B. P. y Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall/CRC.
- Casella, G. y Berger, R. L. (2002). *Statistical Inference*, tomo 2. Duxbury Pacific Grove, CA, 2a edición.
- Casella, G. y George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46, 3, 167–174.
- Chib, S. y Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 4, 327–335.

## Referencias

---

- Craiu, R. V. y Rosenthal, J. S. (2014). Bayesian Computation Via Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 1, 1, 179–201.
- Di Ciccio, T. J. y Monti, A. C. (2011). Inferential aspects of the skew t-distribution. *Quaderni di Statistica*, 13.
- Eastman, J. y Bain, L. J. (1973). A property of maximum likelihood estimators in the presence of location-scale nuisance parameters. *Communications in Statistics*, 2, 1, 23–28.
- Fernández, C. y Steel, M. F. J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93, 441, 359–371.
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 4, 507–521.
- Gauss, F. C. (1809). *Theoria motus corporum coelestium en la sección conicis solem ambientium*. Hamburgo.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. y Rubin, D. (2014). *Bayesian Data Analysis*. CRC Press, Taylor & Francis Group.
- Geman, S. y Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6, 6, 721–741.
- Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics*, 37, 4, 359–363.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 1, 97–109.
- Hernández, A. J. y Zúñiga, R. J. (2013). *Modelos Económicos para el Análisis Económico*. ESIC EDITORIAL, 1a edición.
- Jones, M. C. y Faddy, M. J. (2003). A Skew Extension of the t-Distribution, with Applications. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65, 1, 159–174.
- Kim, H.-J. (2005). Bayesian Estimation for Skew Normal Distributions Using Data Augmentation. *Communications for Statistical Applications and Methods*, 12, 2, 323–333.
- Liseo, B. y Parisi, A. (2013). Bayesian inference for the multivariate skew-normal model: A population Monte Carlo approach. *Computational Statistics & Data Analysis*, 63, 125–138.
- Lorenzo, R. A. y Ruiz, P. A. (2014). A TGARCH Model with an Asymmetric Student's t Distribution and the Rationality Hypotheses of Stock Investors in Latin America. *Munich Personal RePEc Archive*, 20, 1–33.
- M. Legendre, A. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris.
- Maghami, M. y Bahrami, M. (2015). Goodness Of Fit Test For The Skew-t Distribution. *Journal of Mathematics and Computer Science*, 14, 274–283.

## Referencias

---

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. y Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 6, 1087–1092.
- Montgomery, C. D., Peck, A. E. y Vining, G. G. (2006). *Introducción al Análisis de Regresión Lineal*. Compañía editorial Continental, 3a edición.
- Mood, M. A. (1974). *Introduction to the theory of statistics*. McGraw-Hill, 3a edición.
- Naranjo, A. L., Pérez, S. C. J. y Martín, J. J. (2017). Análisis Bayesiano de algunos modelos que usan la distribución t de student asimétrica. En *Aportaciones Recientes a la Estadística en México*, 134–140. AME.
- Parisi, A. y Liseo, B. (2018). Objective Bayesian analysis for the multivariate skew-t model. *Statistical Methods & Applications*, 27, 2, 277–295.
- Pérez-Rodríguez, P., Acosta, P. R., Pérez, E. S., Cruz, C. V., Suárez, E. J. y Crossa, J. (2018). A Bayesian Genomic Regression Model with Skew Normal Random Errors. *G3: Genes, Genomes, Genetics*, 8, 5, 1771–1785.
- Pérez-Rodríguez, P., Villaseñor, A. J. A., Pérez, E. S. y Suárez, E. J. (2017). Bayesian Estimation for the Centered Parameterization of the Skew-Normal Distribution. *Revista Colombiana de Estadística*, 40, 1, 123–140.
- Pérez-Rodríguez, P. y Villaseñor, A. J. A. (2010). On testing the skew normal hypothesis. *Journal of Statistical Planning and Inference*, 140, 11, 3148–3159.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Riascos, V. A. J. (2009). *Métodos Matemáticos y Computacionales en Macroeconomía*. Universidad De Los Andes.
- Rusell-Noriega, M. G. y González-Farías, G. (2002). Análisis de regresión lineal con errores distribuidos normal sesgados. Inf. Téc. I-02-30, Centro de Investigaciones en Matemáticas, A. C.
- Smith, A. y Roberts, G. (1993). Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*, 55, 1, 3–23.
- Sorensen, D. y Gianola, D. (2007). *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. Springer Science & Business Media.
- Student (1908). The Probable Error of a Mean. *Biometrika*, 6, 1, 1–25.
- Szretter, N. M. E. (2017). Regresión lineal. Inf. téc., Universidad de Buenos Aires.
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics.
- Tanner, M. A. y Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82, 398, 528–540.

## Referencias

---

Taylor, J. (2018). *hett: Heteroscedastic t-Regression*. R package version 0.3-2.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Ann. Statist.*, 22, 4, 1701–1728.

Zhu, D. y Galbraith, J. W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics*, 157, 2, 297–305.

# Anexos

## Anexo A: Código en *R* para ajustar el modelo de regresión Bayesiana con errores aleatorios *t-Student*.

El siguiente código en el paquete estadístico *R* (R Core Team, 2018) presenta la función `lm.t()` la cual permite obtener estimaciones de Bayes, utilizando el muestreador de Gibbs, de los parámetros del modelo de regresión con errores aleatorios *t-Student* del capítulo (2).

```
1 #####
2 ####Auxiliary functions
3 #####
4
5 log_L=function(y,mu,sigma2,k)
6 {
7   n=length(y)
8   tmp1=gamma((k+1)/2)/(gamma(k/2)*sqrt(k*pi*sigma2))
9   tmp2=1+(y-mu)^2/(k*sigma2)
10  ans=n*log(tmp1)-0.5*(k+1)*sum(log(tmp2))
11  return(ans)
12 }
13
14 moda=function(z)
15 {
16   dens=density(z)
17   dens$x[dens$y == max(dens$y)] # gives the mode
18 }
19
20 #####
21 ####End of auxiliary functions
22 #####
23
24 #####
25 ####Gibbs
26 #####
27
28 #a-priori beta0=N(0,sigma2_beta)
29 #a-priori Beta=NM(0,sigma2_beta*I)
30 #a-priori sigma^2=1/sigma^2
31
```

## Anexos

---

```
32 ###lm_t is a function that takes as arguments:
33
34 #y: response data with distribution  $y \sim t(\mu, \sigma^2, k)$ .
35 #X: matrix whose columns are the explanatory variables.
36 #k: degrees of freedom.
37 #B: number of gibbs samples.
38 #burnIn: burnIn samples.
39 #thin: number of samples discarded.
40
41 #Initials
42 #beta0_old=0
43 #betaj_old=0 initial value for betaj, j=1,...,p
44 #sigma2_beta=1000000 scale parameter of the prior density of the beta0 and
45 #betaj, j=1,...,p
46 #sigma2=var(y)
47
48 ###lm_t is a function that returns:
49
50 #beta0_mcmc: vector with gibbs sequences of the beta0 parameter.
51 #BETAj_mcmc: matrix whose columns are gibbs sequences of the
52 #beta1,...,betap parameters.
53 #sigma2_mcmc: vector with gibbs sequences of the sigma2 parameter.
54 #fit: list with punctual estimators of the parameters of the model
55 #and statistics such as AIC.
56
57 lm_t=function(y,X,k,B=10000,burnIn=5000,thin=10)
58 {
59   #Initializing
60   n=length(y) #Number of observations
61   p=(ncol(X)+1) #Number of regression coefficients
62   beta0_old=0
63   betaj_old=0
64   sigma2_beta=1000000
65   betaj_oldv=rep(betaj_old,(p-1)) #Initials for beta1,...,betap
66   sigma2=var(y) #Initial for sigma2
67
68   #Objects to store the samples
69   w_matriz=matrix(NA,nrow=B,ncol=n)
70   beta0_mcmc=rep(NA,B)
71   BETAj_matriz=matrix(NA,nrow=B,ncol=(p-1))
72   sigma2_mcmc=rep(NA,B)
73
74   #Errors
75   beta=rep(betaj_old,p)
76   e=y-cbind(1,X)%*%beta
77
78   pb=txtProgressBar(min=1,max=B,style=3)
79
80   for(h in 1:B)
81   {
82     #cat("h=",h,"\n")
83
84     lambda=(sigma2/sigma2_beta)
85
```

## Anexos

---

```
86 #####
87 #Sample from latents w
88 #####
89
90 vm=e^2/sigma2
91 w=rgamma(n,shape=(k+1)/2,rate=(k+vm)/2)
92 w_matriz[h,]=w
93
94 #####
95 #Sample from beta0
96 #####
97
98 t=e+beta0_old
99 swt=sum(w*t)
100 sw=sum(w)
101 beta0_new=rnorm(1,swt/(sw+lambda),sqrt(sigma2/(sw+lambda)))
102 beta0_mcmc[h]=beta0_new
103 delta=beta0_old-beta0_new
104 e=e+delta
105 beta0_old=beta0_new
106
107 #####
108 #Sample from beta1,...,betap
109 #####
110
111 betaj_mcmc=rep(NA,(p-1))
112 for(j in 1:(p-1))
113 {
114   r=e+betaj_oldv[j]*X[,j]
115   swrxj=sum(w*r*X[,j])
116   swxj2=sum(w*X[,j]^2)
117   betaj_new=rnorm(1,(swrxj)/(swxj2+lambda),sqrt(sigma2/(swxj2+lambda)))
118   betaj_mcmc[j]=betaj_new
119   delta=betaj_oldv[j]-betaj_new
120   e=e+delta*X[,j]
121   betaj_oldv[j]=betaj_new
122 }
123 BETAj_matriz[h,]=betaj_mcmc
124
125 #####
126 #Sample from sigma2
127 #####
128
129 SCE=sum(w*e^2)
130 sigma2=1/rgamma(1,shape=n/2,rate=SCE/2)
131 sigma2_mcmc[h]=sigma2
132 setTxtProgressBar(pb,h)
133 }
134
135 close(pb)
136
137 #BurnIn and Thin for w
138 w_mcmc=matrix(nrow=((B-burnIn)*(thin/100)),ncol=n)
139 for(o in 1:n)
```

```

140 {
141   columna=w_matriz[,o][-c(1:burnIn)]
142   columna=columna[seq(1,burnIn,by=thin)]
143   w_mcmc[,o]=columna
144 }
145
146 #BurnIn and Thin for beta0
147 beta0_mcmc=beta0_mcmc[-c(1:burnIn)]
148 beta0_mcmc=beta0_mcmc[seq(1,burnIn,by=thin)]
149
150 #BurnIn and Thin for beta_j, j=1,...,p
151 BETAj_mcmc=matrix(nrow=((B-burnIn)*(thin/100)),ncol=(p-1))
152 for(s in 1:(p-1))
153 {
154   col=BETAj_matriz[,s][-c(1:burnIn)]
155   col=col[seq(1,burnIn,by=thin)]
156   BETAj_mcmc[,s]=col
157 }
158
159 #BurnIn and Thin for sigma2
160 sigma2_mcmc=sigma2_mcmc[-c(1:burnIn)]
161 sigma2_mcmc=sigma2_mcmc[seq(1,burnIn,by=thin)]
162
163 #Point estimators
164
165 #Mean
166 beta0_m=mean(beta0_mcmc)
167 betas_m=rep(NA,(p-1))
168 for(j in 1:(p-1))
169 {
170   media=mean(BETAj_mcmc[,j])
171   betas_m[j]=media
172 }
173 sigma2_m=mean(sigma2_mcmc)
174
175 mds=c(beta0_m,betas_m,sigma2_m)
176 num=length(mds)
177 PostMeans=matrix(nrow=1,ncol=num)
178 for(i in 1:num)
179 {
180   PostMeans[1,i]=mds[i]
181 }
182 bb=length(betas_m)
183 colnames(PostMeans)=c("Beta0",paste("Beta",1:bb,sep=""),"Sigma2")
184 rownames(PostMeans)=c("Means")
185
186 #Mode
187 beta0_md=moda(beta0_mcmc)
188 betas_md=rep(NA,(p-1))
189 for(j in 1:(p-1))
190 {
191   mode=moda(BETAj_mcmc[,j])
192   betas_md[j]=mode
193 }

```

## Anexos

---

```
194 sigma2_md=moda(sigma2_mcmc)
195
196 mos=c(beta0_md,betas_md,sigma2_md)
197 num=length(mos)
198 PostModes=matrix(nrow=1,ncol=num)
199 for(i in 1:num)
200 {
201   PostModes[1,i]=mos[i]
202 }
203 colnames(PostModes)=c("Beta0",paste("Beta",1:bb,sep=""),"Sigma2")
204 rownames(PostModes)=c("Modes")
205
206 #LogLikAtPostMean
207
208 logLikAtPostMean=log_L(y,beta0_m+as.vector(X%*%betas_m),sigma2_m,k)
209
210 KK=ncol(BETAj_mcmc)+3
211 AIC=2*KK-2*logLikAtPostMean
212
213 fit=list()
214 fit$PostMeans=PostMeans
215 fit$PostModes=PostModes
216 fit$logLikAtPostMean=logLikAtPostMean
217 fit$AIC=AIC
218
219 return(list(beta0_mcmc=beta0_mcmc,BETAj_mcmc=BETAj_mcmc,sigma2_mcmc=sigma2_mcmc,w_
    mcmc=w_mcmc,fit=fit))
220 }
221
222 #####
223 #DATA SIMULATION 1
224 #####
225
226 if(FALSE) ### Change (FALSE) to (TRUE) in order to run the example ###
227 {
228   set.seed(436)
229   n=1000 #Sample size.
230   k=100 #Distribution t becomes normal when k->oo.
231   w=rgamma(n,k/2,k/2) #The weights are simulated as 1
232   p=4 #Number of regression parameters.
233   X=matrix(nrow=n,ncol=p-1) #The covariates are fixed.
234   X[,1]=runif(n)
235   X[,2]=runif(n)
236   X[,3]=runif(n)
237   y=rnorm(n,3+1.5*X[,1]+0.3*X[,2]+2*X[,3],1/sqrt(w)) #The answer is simulated, in this
    case the t-Student converges to the Normal.
238   hist(y,col="gold")
239   plot(density(y),main="Density of y")
240
241   #Graph previous important results of the data
242   mean(w) #Show that the weights are 1.
243   par(mfrow=c(2,1))
244   hist(w,col="green")
245   plot(density(w))
```

```

246
247 #Frequentist approach
248 ajustado=lm(y~X[,1]+X[,2]+X[,3])
249 summary(ajustado)
250 anova(ajustado)
251 AIC(ajustado)
252
253 #Call to the function lm_t
254 out=lm_t(y,X,k)
255 #MCMC distributions of parameters
256 beta0_mcmc=out$beta0_mcmc
257 BETAj_mcmc=out$BETAj_mcmc
258 sigma2_mcmc=out$sigma2_mcmc
259 w_mcmc=out$w_mcmc
260 #Point estimators and AIC
261 out$fit
262
263 #Evidence that subsequent distributions by gibbs are correct
264 par(mfrow=c(3,2))
265 hist(w_mcmc[55,],col="green") #Random iteration of the sample of weights.
266 hist(beta0_mcmc,col="cadetblue1")
267 hist(BETAj_mcmc[,1],col="gold")
268 hist(BETAj_mcmc[,2],col="darkmagenta")
269 hist(BETAj_mcmc[,3],col="deeppink")
270 hist(sigma2_mcmc,col="red")
271
272 #Posterior densities graphs
273 par(mfrow=c(3,2))
274 plot(w_mcmc[6,],type="b",ylab=expression(w[i]))
275 plot(density(w_mcmc),xlab=expression(w[i]),main="")
276 plot(beta0_mcmc,type="b",ylab=expression(beta[0]))
277 plot(density(beta0_mcmc),xlab=expression(beta[0]),main="")
278 plot(BETAj_mcmc[,1],type="b",ylab=expression(beta[1]))
279 plot(density(BETAj_mcmc[,1]),xlab=expression(beta[1]),main="")
280
281 par(mfrow=c(3,2))
282 plot(BETAj_mcmc[,2],type="b",ylab=expression(beta[2]))
283 plot(density(BETAj_mcmc[,2]),xlab=expression(beta[2]),main="")
284 plot(BETAj_mcmc[,3],type="b",ylab=expression(beta[3]))
285 plot(density(BETAj_mcmc[,3]),xlab=expression(beta[3]),main="")
286 plot(sigma2_mcmc,type="b",ylab=expression(sigma^2))
287 plot(density(sigma2_mcmc),xlab=expression(sigma^2),main="")
288 }
289
290 #####
291 #DATA SIMULATION 2
292 #####
293
294 if(FALSE) ### Change (FALSE) to (TRUE) in order to run the example ###
295 {
296   n=1000
297   k=3
298   p=3
299   w=rgamma(n,k/2,k/2)

```

## Anexos

---

```
300 X=matrix(nrow=n,ncol=p-1)
301 X[,1]=runif(n)
302 X[,2]=runif(n)
303 y=rnorm(n,4-0.2*X[,1]+1.7*X[,2],1/sqrt(w))
304 hist(y,col="gold")
305 plot(density(y),main="Density of y")
306
307 #Graph previous important results of the data
308 mean(w)
309 par(mfrow=c(2,1))
310 hist(w,col="green")
311 plot(density(w))
312
313 #Frequentist approach
314 ajustado=lm(y~X[,1]+X[,2])
315 summary(ajustado)
316 anova(ajustado)
317 AIC(ajustado)
318
319 #Call to the function lm_t
320 out=lm_t(y,X,k)
321 #MCMC distributions of parameters
322 beta0_mcmc=out$beta0_mcmc
323 BETAj_mcmc=out$BETAj_mcmc
324 sigma2_mcmc=out$sigma2_mcmc
325 w_mcmc=out$w_mcmc
326 #Point estimators and AIC
327 out$fit
328 }
```

---

## Anexo B: Código en *R* para ajustar el modelo de regresión Bayesiana con errores aleatorios *SN* (*BSN*).

El siguiente código de programación en *R* (R Core Team, 2018) presenta la función `lm.sn()` la cual es utilizada para hacer el ajuste de un modelo de regresión Bayesiana cuyos errores se distribuyen de acuerdo a una *SN*, modelo desarrollado en el capítulo (2). El presente está basado en el código tomado de Pérez-Rodríguez *et al.* (2018) disponible en el archivo `.zip` en el sitio <http://www.g3journal.org/content/8/5/1771.supplemental>, y al cual se le realizaron algunos cambios acorde a las necesidades de este trabajo de investigación.

```

1 #####
2 #Load required libraries
3 #####
4 library(truncnorm)
5 library(sn)
6 library(MASS)
7
8 #####
9 ###Auxiliary functions
10 #####
11
12 log_L=function(y,mu,sigma,rho)
13 {
14   EU=sqrt(2/pi)*rho
15   VU=1-2/pi*rho^2
16   SU=sqrt(VU)
17   ystar=y-mu+sigma*EU/SU
18   c=log(2)-1/2*log(2*pi)+log(SU)-log(sigma)
19   n=length(y)
20   tmp1=0.5*(log(pi-2*rho^2)-log(pi*(1-rho^2)))
21   tmp2=exp(tmp1)
22   n*c-1/2*VU/sigma^2*sum(ystar^2)+sum(pnorm(q=rho*tmp2*ystar/sigma,mean = 0, sd = 1,
23     lower.tail = TRUE, log.p = TRUE))
24 }
25 rsn=function(n,rho)
26 {
27   m=c(0,0)
28   V=matrix(c(1,rho,rho,1),nrow=2,ncol=2)
29   d=mvrnorm(n=n,mu=m,Sigma=V)
30   ifelse(d[,1]>0,d[,2],-d[,2])
31 }
32
33 rho_to_lambda=function(rho)
34 {
35   rho/sqrt(1-rho^2)
36 }
37
38 lambda_to_rho=function(lambda)
39 {

```

## Anexos

---

```
40  lambda/sqrt(1+lambda^2)
41  }
42
43  moda=function(z)
44  {
45    dens=density(z)
46    dens$x[dens$y == max(dens$y)] # gives the mode
47  }
48
49  #Function used for the step of metropolis sampling of sigma2
50  #eta1=log(sigma^2)
51  log_eta1=function(eta1,n,z,error,SU,EU,rho)
52  {
53    sigma2=exp(eta1)
54    sigma=sqrt(sigma2)
55    delta=-sigma*rho*z/SU+sigma*EU/SU
56    tmp1=-(n/2+1)*log(sigma2)
57    tmp2=-SU^2/(2*(1-rho^2)*sigma2)*sum((error+delta)^2)
58    tmp3=log(sigma2)
59    ans=tmp1+tmp2+tmp3
60    return(-ans)
61  }
62
63  #Function used for the step of metropolis sampling of rho
64  #eta2=0.5*log((1+rho)/(1-rho))
65  log_eta2=function(eta2,n,z,sigma,error)
66  {
67    sigma2=sigma^2
68    rho=(exp(2*eta2)-1)/(1+exp(2*eta2))
69    EU=sqrt(2/pi)*rho
70    VU=1-2/pi*rho^2
71    SU=sqrt(VU)
72    delta=-sigma*rho*z/SU+sigma*EU/SU
73    tmp1=-VU/(2*(1-rho^2)*sigma2)*sum((error+delta)^2)
74    tmp2=-n/2*log(1-rho^2)
75    tmp3=n*log(SU)
76    tmp4=2*eta2-2*log(1+exp(2*eta2))
77    ans=tmp1+tmp2+tmp3+tmp4
78    return(-ans)
79  }
80
81  #####
82  #### End of auxiliary functions
83  #####
84
85  #NAs not allowed
86  lm_sn=function(y,X,B=5000,burnIn=2500, thin=10,
87                R2=0.5,v1=0.04,v2=0.04)
88  {
89
90    #Fit the model y=beta0+e, where e~SN(0,sigma,gamma_1)
91    fit_sn=selm(y~1)
92    cp=as.vector(fit_sn@param$cp)
93    dp=as.vector(fit_sn@param$dp)
```

```

94
95 beta0=cp[1]
96 sigma=cp[2]
97 lambda=dp[3]
98 rho=lambda_to_rho(lambda)
99
100 EU=sqrt(2/pi)*rho
101 VU=1-2/pi*rho^2
102 SU=sqrt(VU)
103
104 p=ncol(X)
105 n=length(y)
106
107 x2=colSums(X*X)
108 sx=colSums(X)
109
110 beta=rnorm(n=ncol(X),0,sd=0.1)
111 z=abs(rnorm(n))
112
113 sigma2beta=1000000
114
115 error=as.vector(y-X%*%beta)
116 error=error-beta0
117
118 #Objects to store the samples
119 Betas=matrix(NA,nrow=B,ncol=ncol(X))
120 Sigma=rep(NA,B)
121 Rho=rep(NA,B)
122 Beta0=rep(NA,B)
123
124 accept_rate_sigma2=0
125 accept_rate_rho=0
126
127 Xv=as.vector(X)
128
129 nSums=0
130 post_logLik=0
131
132 #####
133 ###Gibbs and Metropolis
134 #####
135
136 #y: response variable
137 #X: incidence matrix
138 #nIter: number of iterations
139 #burnIn: number of samples discarded
140 #R2: R-squared of the model
141
142 #B=5000
143 #burnIn=B/2
144
145 #R2=0.5
146
147 for(iter in 1:B)

```

## Anexos

---

```
148 {
149   #cat("iter=",iter,"beta0=",round(beta0,4),"beta1=",round(beta,4),"sigma=",round(
150     sigma,4),"rho=",round(rho,4),"\\n")
151   #####
152   #Sample from beta0
153   #####
154
155   delta=-sigma*rho*z/SU+sigma*EU/SU
156
157   yc=error+beta0+delta
158   c1=n+(1-rho^2)*sigma^2/(SU^2*sigma2beta)
159   c2=sum(yc)
160   m=c2/c1
161   v=(1-rho^2)*sigma^2/(SU^2*c1)
162
163   beta0_old=beta0
164   beta0_new=rnorm(1,m,sqrt(v))
165   beta0=beta0_new
166
167   #Update the error
168   error=error+(beta0_old-beta0_new)
169   Beta0[iter]=beta0
170
171   #####
172   #Sample from beta1,...,betap
173   #####
174
175   for(j in 1:p)
176   {
177     yc=error+X[,j]*beta[j]+delta
178     c1=x2[j]+(1-rho^2)*sigma^2/(SU^2*sigma2beta)
179     c2=sum(yc*X[,j])
180     m=c2/c1
181     v=(1-rho^2)*sigma^2/(SU^2*c1)
182
183     bold=beta[j]
184     bnew=rnorm(1,m,sqrt(v))
185     beta[j]=bnew
186
187     #Update the error
188     error=error+(bold-bnew)*X[,j]
189   }
190   Betas[iter,]=beta
191
192   #####
193   #Sample from z
194   #####
195
196   z=rtruncnorm(n=n,a=0, b=Inf, mean = rho*((error/sigma)*SU+EU), sd = sqrt(1-rho^2))
197
198   #####
199   #Sample from sigma^2
200   #####
```

```

201
202   eta1=log(sigma^2)
203   eta1p=eta1+rnorm(n=1,mean=0,sd=sqrt(v1))
204
205   U_eta1=log_eta1(eta1,n,z,error,SU,EU,rho)
206   U_eta1p=log_eta1(eta1p,n,z,error,SU,EU,rho)
207
208   accept = (U_eta1-U_eta1p) > log(runif(1))
209
210   if(accept)
211   {
212     sigma2=exp(eta1p)
213     sigma=sqrt(sigma2)
214     accept_rate_sigma2=accept_rate_sigma2+1
215   }
216
217     Sigma[iter]=sigma
218
219     #####
220     #Sample from rho
221     #####
222
223     eta2=0.5*log((1+rho)/(1-rho))
224     eta2p=eta2+rnorm(n=1,mean=0,sd=sqrt(v2))
225
226     U_eta2=log_eta2(eta2,n,z,sigma,error)
227     U_eta2p=log_eta2(eta2p,n,z,sigma,error)
228
229     accept = (U_eta2-U_eta2p) > log(runif(1))
230
231     if (accept) {
232       rho=(exp(2*eta2p)-1)/(1+exp(2*eta2p))
233       accept_rate_rho=accept_rate_rho+1;
234     }
235
236     EU=sqrt(2/pi)*rho
237     VU=1-2/pi*rho^2
238     SU=sqrt(VU)
239     Rho[iter]=rho
240
241
242     if(iter>burnIn & iter%%thin == 0)
243     {
244       #mu=y-error
245       nSums=nSums+1
246       k=(nSums-1)/nSums
247       logLik=log_L(y,y-error,sigma,rho)
248       post_logLik = post_logLik * k + logLik/nSums
249     }
250
251   }
252
253   #Burn-in and thin
254

```

## Anexos

---

```
255 Beta0=Beta0[(burnIn+1):B]
256 index=seq(1,length(Beta0),by=thin)
257 Beta0=Beta0[index]
258
259 filas=(B-burnIn)*(thin/100)
260 Betasj=matrix(nrow=filas,ncol=p)
261 for(j in 1:p)
262 {
263   col=Betas[,j][(burnIn+1):B]
264   col=col[index]
265   Betasj[,j]=col
266 }
267
268 Sigma=Sigma[(burnIn+1):B]
269 Sigma=Sigma[index]
270 Rho=Rho[(burnIn+1):B]
271 Rho=Rho[index]
272 rm(index)
273
274 #Point estimators
275
276 #Mean
277 beta0_m=mean(Beta0)
278 betas_m=rep(NA,p)
279 for(j in 1:p)
280 {
281   media=mean(Betasj[,j])
282   betas_m[j]=media
283 }
284 sigma_m=mean(Sigma)
285 rho_m=mean(Rho)
286 lambda=rho_to_lambda(Rho)
287 lambda_m=mean(lambda)
288
289 mds=c(beta0_m,betas_m,sigma_m,rho_m,lambda_m)
290 num=length(mds)
291 PostMeans=matrix(nrow=1,ncol=num)
292 for(i in 1:num)
293 {
294   PostMeans[1,i]=mds[i]
295 }
296 bb=length(betas_m)
297 colnames(PostMeans)=c("Beta0",paste("Beta",1:bb,sep=""),"Sigma","Rho","Lambda")
298 rownames(PostMeans)=c("Means")
299
300 #Mode
301 beta0_md=moda(Beta0)
302 betas_md=rep(NA,p)
303 for(j in 1:p)
304 {
305   mode=moda(Betasj[,j])
306   betas_md[j]=mode
307 }
308 sigma_m=moda(Sigma)
```

## Anexos

---

```
309 rho_m=moda(Rho)
310 lambda_m=moda(lambda)
311
312 mos=c(beta0_m,betas_m,sigma_m,rho_m,lambda_m)
313 num=length(mos)
314 PostModes=matrix(nrow=1,ncol=num)
315 for(i in 1:num)
316 {
317   PostModes[1,i]=mos[i]
318 }
319 colnames(PostModes)=c("Beta0",paste("Beta",1:bb,sep=""),"Sigma","Rho","Lambda")
320 rownames(PostModes)=c("Modes")
321
322 #LogLikAtPostMean
323
324 logLikAtPostMean=log_L(y,beta0_m+as.vector(X%*%betas_m),sigma_m,rho_m)
325
326 K=K+ncol(Betas)+3
327 AIC=2*K-2*logLikAtPostMean
328 pD = -2 * (post_logLik - logLikAtPostMean)
329 DIC = pD - 2 * post_logLik
330
331 fit=list()
332 fit$PostMeans=PostMeans
333 fit$PostModes=PostModes
334 fit$logLikAtPostMean=logLikAtPostMean
335 fit$postMeanLogLik=post_logLik
336 fit$AIC=AIC
337 fit$pD=pD
338 fit$DIC=DIC
339
340 acceptance_rates_RWM=list()
341 acceptance_rates_RWM$accept_rate_rho=accept_rate_rho/B
342 acceptance_rates_RWM$accept_rate_sigma=accept_rate_sigma2/B
343
344 return(list(beta0=Beta0,betas=Betasj,sigma=Sigma,rho=Rho,lambda=lambda,fit=fit,
345           acceptance_rates_RWM=acceptance_rates_RWM))
}
```

---

## Anexo C: Código en *R* para ajustar el modelo de regresión Bayesiana con errores aleatorios *ST* de [Azzalini y Capitanio \(2003\)](#), (*BST*).

El siguiente código de programación en *R* ([R Core Team, 2018](#)) presenta la función `lm_st()` la cual es utilizada para hacer el ajuste de un modelo de regresión Bayesiana cuyos errores se distribuyen de acuerdo a una *ST* de [Azzalini y Capitanio \(2003\)](#), modelo desarrollado en el capítulo (3) y cuya aplicación se ilustra en el capítulo (4) para el conjunto de datos de la compañía Martin Marietta.

```

1 #####
2 #Load requiered libraries
3 #####
4
5 library(truncnorm)
6 library(MASS)
7
8 #####
9 #####Auxiliary functions
10 #####
11
12 log_L=function(y,xi,omega,lambda,k)
13 {
14   z=(y-xi)/omega
15   py=2/omega*dt(z,df=k)*pt(lambda*z*sqrt((k+1)/(k+z^2)),df=k+1)
16   ans=sum(log(py))
17 }
18
19 beta0_cp=function(beta0,rho,omega,k)
20 {
21   bv=(k/pi)^(1/2)*(gamma((k-1)/2)/gamma(k/2))
22   ans=beta0+rho*bv*omega
23 }
24
25 moda=function(z)
26 {
27   dens=density(z)
28   dens$x[dens$y == max(dens$y)] # gives the mode
29 }
30
31 rsn=function(n,rho)
32 {
33   m=c(0,0)
34   V=matrix(c(1,rho,rho,1),nrow=2,ncol=2)
35   d=mvrnorm(n=n,mu=m,Sigma=V)
36   ifelse(d[,1]>0,d[,2],-d[,2])
37 }
38
39 rst=function(n,rho,k)
40 {
41   u=rsn(n,rho)

```

## Anexos

---

```
42 v=rchisq(n,df=k,ncp=0)
43 return(u/sqrt(v/k))
44 }
45
46 rho_to_lambda=function(rho)
47 {
48   rho/sqrt(1-rho^2)
49 }
50
51 lambda_to_rho=function(lambda)
52 {
53   lambda/sqrt(1+lambda^2)
54 }
55
56 #Function used for the step of Metropolis sampling of omega2
57 #eta1=log(omega^2)
58 log_eta1=function(eta1,n,omega,rho,error,z,v,k)
59 {
60   omega2=exp(eta1)
61   omega=sqrt(omega2)
62   tmp1=-(n/2+1)*eta1
63   tmp2=-1/(2*(1-rho^2))*sum((error/omega*sqrt(v/k)-rho*z)^2)
64   tmp3=log(omega2)
65   ans=tmp1+tmp2+tmp3
66   return(-ans)
67 }
68
69 #Function used for the step of Metropolis sampling of rho
70 #eta2=0.5*log((1+rho)/(1-rho))
71 log_eta2=function(eta2,n,omega,error,z,v,k)
72 {
73   rho=(exp(2*eta2)-1)/(1+exp(2*eta2))
74   tmp1=-n/2*log(1-rho^2)
75   tmp2=-1/(2*(1-rho^2))*sum((error/omega*sqrt(v/k)-rho*z)^2)
76   tmp3=2*eta2-2*log(1+exp(2*eta2))
77   ans=tmp1+tmp2+tmp3
78   return(-ans)
79 }
80
81 #log of Density function of conditional for latent variable v
82 log_f_v=function(v,A,B,k)
83 {
84   (((k+1)/2)-1)*log(v)-A*v-B*sqrt(v)
85 }
86
87 #log of density function of square of a gamma random variable with
88 #shape and rate parameters
89 #for Algorithm of Acceptance and Rejection of the latent variable v
90 log_f_g2=function(v,shape,rate)
91 {
92   shape*log(rate)+(shape/2-1)*log(v)-rate*sqrt(v)-log(2)-lgamma(shape)
93 }
94
95 #####
```

## Anexos

---

```
96 #### End of auxiliary functions
97 #####
98
99 #####
100 ####Gibbs
101 #####
102
103 #a-priori beta0=N(0,sigma2_beta)
104 #a-priori Beta=NM(0,sigma2_beta*I)
105 #a-priori omega^2=1/omega^2
106
107 ###lm_t is a function that takes as arguments:
108
109 #y: response data with distribution  $y \sim ST(x_i, \omega^2, \lambda, k)$ .
110 #X: matrix whose columns are the explanatory variables.
111 #B: number of gibbs samples.
112 #burnIn: burnIn samples.
113 #thin: number of samples discarded.
114
115 #Initials
116 #k: Degrees of freedom
117 #sigma2_beta=1*10^6 scale parameter of the prior density
118 #of the beta0 and betaj, j=1,...,p
119
120 #The initial values of the following parameters are calculated within the function
121 #fm=lm(y~x) Fit the model  $y = \beta_0 + e$ , where  $e \sim N(0, \sigma^2)$ 
122 #beta=coef(fm)
123 #beta0=beta[1]
124 #beta1=beta[2]
125 #omega2=sum(residuals(fm)^2)/(n-2)
126 #omega=sqrt(omega2)
127 #v=rchisq(n,k)
128 #z=abs(rnorm(n))
129 #rho=0.5
130
131 ###lm_t is a function that returns:
132
133 #Beta0: vector with gibbs sequences of the beta0 parameter.
134 #Betasj: matrix whose columns are gibbs sequences of the
135 #beta1,...,betap parameters.
136 #Omega: vector with gibbs sequences of the omega parameter.
137 #rhos: vector with gibbs sequences of the rho parameter.
138 #fit: list with punctual estimators of the parameters of the model
139 #and statistics such as AIC.
140
141 #NAs not allowed
142
143 lm_st=function(y,X,k,B=10000,burnIn=5000,thin=10)
144 {
145
146 #Initializing
147 n=length(y)
148 p=ncol(X)
149 fm=lm(y~X) #Fit the model  $y = \beta_0 + e$ , where  $e \sim N(0, \sigma)$ 
```

## Anexos

---

```
150 beta=coef(fm)
151 beta0=beta[1]
152 betajs=beta[-1]
153 omega2=sum(residuals(fm)^2)/(n-2)
154 omega=sqrt(omega2)
155 sigma2_beta=1e6 #Non informative distribution for betas
156 v=rchisq(n,k)
157 z=abs(rnorm(n))
158 rho=0.5
159
160 #Errors
161 error=as.vector(y-cbind(1,X)%*%beta)
162
163 #Objects to store the samples
164 rhos=rep(NA,B)
165 Beta0=rep(NA,B)
166 Betas=matrix(NA,nrow=B,ncol=p)
167 Omega=rep(NA,B)
168
169 accept_rate_omega=0
170 accept_rate_rho=0
171
172 pb=txtProgressBar(min=1,max=B,style=3)
173
174 for(iter in 1:B)
175 {
176   #cat("iter=",iter,"\n")
177
178   #####
179   #Sample from beta0
180   #####
181
182   c1=sum(v)/k+(1-rho^2)*omega^2/sigma2_beta
183   d=(error+beta0)*sqrt(v/k)-rho*omega*z
184   c2=sum(d*sqrt(v/k))
185   beta0_old=beta0
186   beta0_new=rnorm(1,c2/c1,sqrt((1-rho^2)*omega^2/c1))
187   beta0=beta0_new
188   Beta0[iter]=beta0
189   error=error+(beta0_old-beta0_new)
190
191   #####
192   #Sample from beta1,...,betap
193   #####
194
195   betaj_mcmc=rep(NA,p)
196   for(j in 1:p)
197   {
198     f=(error+betajs[j]*X[,j])*sqrt(v/k)-rho*omega*z
199     c1=sum(X[,j]^2*v)/k+(1-rho^2)*omega^2/sigma2_beta
200     c2=sum(f*X[,j])*sqrt(v/k)
201     betaj_old=betajs[j]
202     betaj_new=rnorm(1,c2/c1,sqrt((1-rho^2)*omega^2/c1))
203     betaj_mcmc[j]=betaj_new
```

```

204     betajs[j]=betaj_new
205     error=error+(betaj_old-betaj_new)*X[,j]
206   }
207   Betas[iter,]=betaj_mcmc
208   #cat("beta0=",beta0,"\n")
209   #cat("beta1=",beta1,"\n")
210
211   #####
212   #Sample from latent variables (Z)
213   #####
214
215   g=error*sqrt(v/k)
216   z=rtruncnorm(n=n,a=0, b=Inf, mean = rho*g/omega, sd = sqrt(1-rho^2))
217
218   #####
219   #Sample from omega
220   #####
221
222   eta1=log(omega^2)
223   eta1p=eta1+rnorm(n=1,mean=0,sd=sqrt(0.04))
224
225   U_eta1=log_eta1(eta1,n,omega,rho,error,z,v,k)
226   U_eta1p=log_eta1(eta1p,n,omega,rho,error,z,v,k)
227
228   accept = (U_eta1-U_eta1p) > log(runif(1))
229
230   if(accept)
231   {
232     omega2=exp(eta1p)
233     omega=sqrt(omega2)
234     accept_rate_omega=accept_rate_omega+1
235   }
236
237   Omega[iter]=omega
238   #cat("omega=",omega,"\n")
239
240   #####
241   #Sample from rho
242   #####
243
244   eta2=0.5*log((1+rho)/(1-rho))
245   eta2p=eta2+rnorm(n=1,mean=0,sd=sqrt(0.04))
246
247   U_eta2=log_eta2(eta2,n,omega,error,z,v,k)
248   U_eta2p=log_eta2(eta2p,n,omega,error,z,v,k)
249
250   accept = (U_eta2-U_eta2p) > log(runif(1))
251
252   if (accept) {
253     rho=(exp(2*eta2p)-1)/(1+exp(2*eta2p))
254     accept_rate_rho=accept_rate_rho+1
255   }
256
257   rhos[iter]=rho

```

```

258   #cat("rho=",rho,"\n")
259
260   #####
261   #Sample from latent variables (V)
262   #####
263
264   b=1/sqrt(k)*error
265   A=(b^2+(1-rho^2)*omega^2)/(2*(1-rho^2)*omega^2)
266   C=(k+1)/2
267   Bi=-2*b*rho*omega*z/((1-rho^2)*omega^2)
268
269   par1=2*C
270   par2=(Bi+sqrt(Bi^2+8*A*(2*C+1)))/2
271
272   for(i in 1:n)
273   {
274     #cat("i=",i,"\n")
275     reject=TRUE
276     v_optimal=((par2[i]-Bi[i])/(2*A[i]))^2
277     lM=log_f_v(v_optimal,A[i],Bi[i],k)-log_f_g2(v_optimal,shape=par1,rate=par2[i])
278     #cat("lM=",lM,"\n")
279     #cat("v_optimal=",v_optimal,"\n")
280     while(reject)
281     {
282       v_proposal=(rgamma(1,shape=par1,rate=par2[i]))^2
283       if(log(runif(1))+lM+log_f_g2(v_proposal,shape=par1,rate=par2[i])<log_f_v(v_
284         proposal,A[i],Bi[i],k))
285       {
286         reject=FALSE
287         v[i]=v_proposal
288       }
289     }
290
291     setTxtProgressBar(pb,iter)
292
293   }
294
295   close(pb)
296
297   #MCMC Complete
298   BetaOMCMC<<-Beta0
299   BetasMCMC<<-Betas
300   OmegaMCMC<<-Omega
301   RhoMCMC<<-rhos
302
303   #Burn-in and thin
304
305   Beta0=Beta0[(burnIn+1):B]
306   index=seq(1,length(Beta0),by=thin)
307   Beta0=Beta0[index]
308
309   filas=(B-burnIn)*(thin/100)
310   Betasj=matrix(nrow=filas,ncol=p)

```

```

311 for(j in 1:p)
312 {
313   col=Betas[,j][(burnIn+1):B]
314   col=col[index]
315   Betasj[,j]=col
316 }
317
318 Omega=Omega[(burnIn+1):B]
319 Omega=Omega[index]
320 rhos=rhos[(burnIn+1):B]
321 rhos=rhos[index]
322 rm(index)
323
324 #Point estimators
325
326 #Mean
327 beta0_m=mean(Beta0)
328 betas_m=rep(NA,p)
329 for(j in 1:p)
330 {
331   media=mean(Betasj[,j])
332   betas_m[j]=media
333 }
334 omega_m=mean(Omega)
335 rho_m=mean(rhos)
336 lambda=rho_to_lambda(rhos)
337 lambda_m=mean(lambda)
338
339 mds=c(beta0_m,betas_m,omega_m,rho_m,lambda_m)
340 num=length(mds)
341 PostMeans=matrix(nrow=1,ncol=num)
342 for(i in 1:num)
343 {
344   PostMeans[1,i]=mds[i]
345 }
346 bb=length(betas_m)
347 colnames(PostMeans)=c("Beta0",paste("Beta",1:bb,sep=""),"Omega","Rho","Lambda")
348 rownames(PostMeans)=c("Means")
349
350 #Mode
351 beta0_md=moda(Beta0)
352 betas_md=rep(NA,p)
353 for(j in 1:p)
354 {
355   mode=moda(Betasj[,j])
356   betas_md[j]=mode
357 }
358 omega_md=moda(Omega)
359 rho_md=moda(rhos)
360 lambda_md=moda(lambda)
361
362 mos=c(beta0_md,betas_md,omega_md,rho_md,lambda_md)
363 num=length(mos)
364 PostModes=matrix(nrow=1,ncol=num)

```

## Anexos

---

```
365 for(i in 1:num)
366 {
367   PostModes[1,i]=mos[i]
368 }
369 colnames(PostModes)=c("Beta0",paste("Beta",1:bb,sep=""), "Omega", "Rho", "Lambda")
370 rownames(PostModes)=c("Modes")
371
372 mu0_cp=beta0_cp(beta0_md,rho_md,omega_md,k)
373
374 #LogLikAtPostMean
375
376 logLikAtPostMean=log_L(y,beta0_m+as.vector(X%*%betas_m),mean(omega_m),lambda_m,k)
377
378 KK=ncol(Betas)+4
379 AIC=2*KK-2*logLikAtPostMean
380
381 fit=list()
382 fit$PostMeans=PostMeans
383 fit$PostModes=PostModes
384 fit$mu0_cp=mu0_cp
385 fit$logLikAtPostMean=logLikAtPostMean
386 fit$AIC=AIC
387
388 acceptance_rates_RWM=list()
389 acceptance_rates_RWM$accept_rate_omega=accept_rate_omega/B
390 acceptance_rates_RWM$accept_rate_rho=accept_rate_rho/B
391
392 return(list(rhos=rhos,Beta0=Beta0,Betasj=Betasj,Omega=Omega,lambda=lambda,acceptance_
   rates_RWM=acceptance_rates_RWM,fit=fit))
393
394 }
395
396 #####
397 #Example of an iteration in the Acceptance and Rejection Algorithm
398 #####
399
400 if(FALSE) ### Change (FALSE) to (TRUE) in order to run the example ###
401 {
402
403 #Example accept reject
404 v_optimal=5.030004
405 A=0.5251403
406 B=0.01207629
407 k=3.31
408 par1=4.31
409 par2=2.367611
410
411 lM=log_f_v(v_optimal,A,B,k)-log_f_g2(v_optimal,shape=par1,rate=par2)
412
413 v=seq(0.01,15,length.out=1000)
414 f1=log_f_v(v,A,B,k)
415 g2=log_f_g2(v,shape=par1,rate=par2)
416
417 plot(v,exp(lM)*exp(g2),type="l",lty=1,xlab="v",ylab="p(v|resto)",lwd=2)
```

```
418 points(v,exp(f1),type="l",lty=2,lwd=2)
419 points(v,exp(g2),type="l",lty=3,lwd=2)
420 legend("topright",lty=1:3,legend=c("Mg(v)", "p(v|resto)", "g(v)"))
421
422 s=0
423 accepted=0
424 rejected=0
425 while(s<1000)
426 {
427   v_proposal=(rgamma(1,shape=par1,rate=par2))^2
428   u=runif(1)
429   if(log(u)+lM+log_f_g2(v_proposal,shape=par1,rate=par2)<log_f_v(v_proposal,A,B,k))
430   {
431     #Accept
432     points(v_proposal,u*exp(lM)*exp(log_f_g2(v_proposal,shape=par1,rate=par2)),col="
green")
433     accepted=accepted+1
434   }else{
435     #Reject
436     points(v_proposal,u*exp(lM)*exp(log_f_g2(v_proposal,shape=par1,rate=par2)),col="
red",pch=4)
437     rejected=rejected+1
438   }
439   s=s+1
440 }
441
442 acceptance_rate=accepted/1000
443 acceptance_rate
444 rejection_rate=rejected/1000
445 rejection_rate
446
447 }
```

---

## Anexo D: Código en *R* para el ajuste del modelo de regresión *Normal*, *t-Student*, *BSN* y *BST* para el conjunto de datos Martin Marietta.

El siguiente código en *R* (R Core Team, 2018) presenta el ajuste del modelo de regresión *Normal* (usando la función `lm()`), *t-Student* (usando la función `lm_t()` del Anexo A), *BSN* (usando la función `lm_sn()` del Anexo B) y *BST* (usando la función `lm_st()` del Anexo C), para el conjunto de datos Martin Marietta que son cargados desde el archivo `DataMarietta.csv` (también disponibles en la librería `hett` de *R*).

```

1 #####
2 # REQUIRED .R FILES IN THE SAME FOLDER
3 #####
4 # RLM_tStudent_nuFijo_Final.R (Anexo A)
5 # fit_sn_centered_pure_R_ALDO.R (Anexo B)
6 # st_ALDO_2.R (Anexo C)
7
8 #####
9 # READ DATA FROM ARCHIVE .CSV
10 #####
11 setwd("~/Documents/TESIS/COMPARACION DE MODELOS")
12 datos=read.csv(file="DataMarietta.csv",header=TRUE)
13 y=datos$M.Marietta
14 X=as.matrix(datos$CRSP)
15
16 #####
17 # REGRESSION WITH ERRORS N(0,sigma2)
18 #####
19 fit_n=lm(y~X)
20 summary(fit_n)
21
22 #####
23 # REGRESSION WITH ERRORS t(0,sigma2,k)
24 #####
25 source("RLM_tStudent_nuFijo_Final.R")
26 fit_t=lm_t(y,X,k=3.00,B=10000,burnIn=5000,thin=10)
27 fit_t
28
29 #####
30 # REGRESSION WITH ERRORS SN(0,sigma2,rho)
31 #####
32 source("fit_sn_centered_pure_R_ALDO.R")
33 fit_sn=lm_sn(y,X,B=10000,burnIn=5000,thin=10,R2=0.5,v1=0.04,v2=0.04)
34 fit_sn
35
36 #####
37 # REGRESSION WITH ERRORS ST(0,omega2,lambda,k)
38 #####
39 source("st_ALDO_2.R")
40 fit_st=lm_st(y,X,k=3.00,B=10000,burnIn=5000,thin=10)
41 fit_st

```