



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

CONSTRUCCIÓN DE UN MODELO
AR-ARCH VÍA VARIABLES
LATENTES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

MIGUEL ANGEL CONTRERAS ORTIZ



DIRECTOR DE TESIS:
DR. ALBERTO CONTRERAS CRISTÁN

2011



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A Dios primeramente ya que Él me ha dado todo y me ha permitido concluir mi carrera al igual que este trabajo, todo ha sido mediante su gran ayuda y misericordia que tiene para con sus hijos. ¡Gracias Papá!

A mis padres, gracias por su apoyo, motivación y comprensión durante todos mis estudios que finalmente pude consumir con este trabajo, sin duda gran parte del mérito fue gracias a su amor y apoyo incondicional. ¡Los amo mucho!

A mis amigos quienes estuvieron conmigo durante toda la carrera, que en los buenos y malos momentos, siempre pudo contar con ellos y aún en este trabajo pude ver su apoyo y compañerismo que tuvieron siempre. Gracias por la amistad que me han brindado y que continúa hasta el momento.

Al Dr. Eduardo Guitérrez Peña, a la Dra. Ruth Fuentes García, a la Mat. Margarita Chavez Cano y al Dr. Luis Rincón Soís. Gracias por sus comentarios y sugerencias que fueron muy importantes en la etapa final de este trabajo y que aportaron mucho valor que enriqueció el contenido de éste, gracias por su tiempo y esfuerzo.

A la UNAM y a la Facultad de Ciencias por la formación académica que me brindaron, sin duda alguna estoy en deuda con Alma Mater, la institución académica más importante del país, que me llena de orgullo rotundamente el haber egresado de esta Universidad, ejemplo de las grandes cosas que tiene nuestro país.

**Un agradecimiento especial para
el Dr. Alberto Contreras Cristán.**

No tengo palabras para agradecer su incesable apoyo durante todo el tiempo que nos llevó terminar este trabajo, gracias por todas y cada una de las asesorías que me brindó en las cuales, sin importar el tiempo que nos llevara, no terminaba de explicarme hasta aclarar mis dudas por completo.

Me llena de mucho orgullo haber podido trabajar a su lado, gracias por la oportunidad que me brindó al aceptar dirigir mi trabajo. Desde un inicio supe que, al elegirlo como mi asesor, aprendería muchísimo de su parte y siendo sincero, lo admiro mucho, no solo por ser un excelente profesor y una persona tan brillante, sino por ser una gran persona que realmente ama lo que hace y da todo por sus alumnos, en verdad mil gracias.

Índice general

Agradecimientos	I
Introducción	v
1. Fundamentos y motivación de las variables latentes	1
1.1. Procesos Estocásticos	1
1.1.1. Cadenas de Markov	2
1.2. Estadística Bayesiana	5
1.2.1. Proceso de aprendizaje	5
1.2.2. Predicción	10
1.2.3. Distribuciones No Informativas	16
1.3. Variables latentes	17
1.4. Algoritmo EM	19
2. Series de Tiempo	27
2.1. Procesos Estacionarios	29
2.1.1. Ruido Blanco	33
2.1.2. Promedios Móviles	34
2.1.3. Autorregresivo	34
2.1.4. Proceso ARMA(p, q)	39
2.1.5. La función de autocovarianza y autocorrelación	47
2.1.6. La función de autocorrelación parcial	49
2.2. Selección de modelos	52
2.2.1. Estimación de parámetros por máxima verosimilitud	52
2.2.2. Análisis de Residuales	54
2.2.3. Predicción	57
2.3. Series de tiempo no lineales	58
2.3.1. Propiedades básicas del proceso ARCH(q)	58

2.3.2.	Propiedades básicas del proceso GARCH(p, q)	64
2.3.3.	Modelo ARMA-GARCH	69
3.	Series de tiempo vía variables latentes	71
3.1.	El modelo de Pitt y Walker	71
3.2.	Observaciones y variables latentes	72
3.3.	Un proceso ARCH con variables latentes	74
3.4.	Construcción del modelo AR-ARCH	76
3.5.	Usando el algoritmo EM	78
3.5.1.	Paso - E	79
3.5.2.	Paso - M	82
3.6.	Simulación y ajuste del modelo	83
3.7.	Aplicación a datos financieros	88
3.7.1.	Predicción para observaciones futuras	100
3.7.2.	Valuación de opciones y volatilidad estocástica	103
3.7.3.	Estimación del VaR	110
4.	Conclusiones	113
A.	Algoritmo Gibbs sampler	115
B.	Teoremas y cálculos	121
B.1.	Teorema de convergencia monótona (TCM)	121
B.2.	Teorema de convergencia dominada (TCD)	121
B.3.	Estimación de parámetros para el proceso ARCH(1)	122
B.4.	Cálculos de la transición del proceso del Capítulo 4	124
B.5.	Cálculos del Paso-E del Capítulo 4	127
C.	Códigos en MATLAB	131
C.1.	Códigos del Capítulo 2	131
C.2.	Códigos del Capítulo 3	132
C.3.	Códigos del Capítulo 4	135
	Bibliografía	151

Introducción

La estadística provee múltiples herramientas que nos permiten modelar el comportamiento de datos pertenecientes a instrumentos financieros como son el análisis multivariado, el análisis de regresión, las series de tiempo, etc. Estas últimas son el objeto central de este trabajo ya que se han desarrollado modelos cuyo interés ha nacido en poder describir el comportamiento de los instrumentos financieros y hacer una predicción del precio de estos mismos. Engle (1982) y Bollerslev (1986) realizaron investigación en los campos de finanzas y econometría creando los conocidos modelos *ARCH* y *GARCH* respectivamente, y que han sido utilizados para el tratamiento de la volatilidad de los activos financieros.

Existen diversas variantes y mejorías a estos modelos, entre ellos el *Exponential GARCH*, *TS-GARCH*, *GJR-GARCH*, *T-ARCH*, *N-ARCH*, *Log-ARCH*, *ARMA-GARCH*, entre otros. Ver Engle (1982) y (1987), Bollerslev (1986) y (1987), Glosten, L.R., R. Jagannathan, and D.E. Runkle (1993), Nelson, D.B. (1991) y Würtz, Chalabi, Luksan (2006). En particular el proceso ARMA-GARCH tiene una gran diferencia con respecto a los otros ya que no comparte el supuesto de que los retornos de un activo financiero tengan una media igual a cero sino que usa la hipótesis de que la media condicional sigue un proceso ARMA, considerando entonces que tanto la media como la varianza se modelan a través de una serie de tiempo.

Por otra parte en el campo de la estadística Bayesiana se han desarrollado diversas metodologías que motivan las ideas que se implementan en esta tesis, como son el *Gibbs Sampler* (o muestreo de Gibbs) del cual se desprende la idea base para incorporar variables latentes o auxiliares en modelos probabilísticos. Ver Pitt, Chatfield, Walker (2002) y Pitt, Walker (2005). Los autores que se acaban de mencionar desarrollaron algunos artículos en los que usan los supuestos de los procesos latentes para construir series de tiempo

estacionarias con una distribución marginal específica para las observaciones, definiendo las distribuciones condicionales para las variables observables y las no observables de manera apropiada.

Las series de tiempo que ellos construyeron usando variables latentes fueron los modelos AR(1), ARCH(1) y GARCH(1,1). El modelo que se propone en esta tesis es un proceso con media AR(1) y varianza ARCH(1) considerando en esta última parte el uso de variables latentes.

Estos modelos tienen la ventaja de poder especificar una distribución marginal de probabilidad para los datos con los que estamos trabajando distinta a la distribución Normal que es con la que se trabaja regularmente. Con mayor detalle, esta idea consiste en “imaginar” la existencia de una serie de tiempo latente que es inherente a la serie de tiempo que sí observamos. Esta serie latente determinará la relación de dependencia estocástica entre dos observaciones consecutivas de la serie observada.

En la construcción del modelo se considera un proceso ARCH(1) utilizando la idea de las variables latentes y metodologías bayesianas al igual que en el artículo de Pitt, Walker (2005) pero con la innovación de introducir una media AR(1) lo cual marca la diferencia entre el modelo de Pitt-Walker y el nuestro. Aunque la idea es la misma hay varias cosas que se tuvieron que desarrollar para este modelo en particular ya que la verosimilitud del modelo se modifica, por ejemplo. Otra de las diferencias es que Pitt y Walker emplean métodos MCMC para la estimación de parámetros y nosotros usamos el *algoritmo EM* usado ampliamente en la comunidad estadística, ya que este método es especial para el análisis de datos en los que hay variables que no pueden ser observadas, por lo cual es ideal para el modelo que se propone. Cabe mencionar que el algoritmo Gibbs Sampler también puede utilizarse para estimar parámetros en modelos que involucran la existencia de variables latentes.

Dentro de las aplicaciones que consideramos, la primera consiste modelar y predecir log-rendimientos de activos financieros. La segunda aplicación consiste en estimar la volatilidad de dichos log-rendimientos, sustituirla en la ecuación de Black-Scholes y de esta manera poder valorar una opción financiera suscrita sobre esta acción, suponiendo que la volatilidad es estocástica.

Por tanto, en este trabajo se expone un metodología para poder construir series de tiempo con una distribución marginal específica para las observaciones a través del ejemplo que desarrollamos, el cual fue un proceso AR(1)-ARCH(1), construido con ayuda de variables latentes y metodologías Bayesianas. Lo segundo que se expone es una aplicación mas en la que puede usarse el algoritmo EM para estimar parámetros en una serie de tiempo que involucre en su estructura una o más variables latentes. Este método, al ser parte de las técnicas motivadas por la estadística Bayesiana junto con los métodos MCMC, nos muestran que otro tipo de inferencia distinta de la manejada en el enfoque clásico, en ocasiones se convierte en una necesidad más que en una alternativa para construir modelos estadísticos diferentes e incluso mejores de los que existen en la actualidad.

Capítulo 1

Fundamentos y motivación de las variables latentes

De manera muy general y elemental, en este primer capítulo describiremos las herramientas de la probabilidad y la estadística que nos servirán para abordar la teoría de las series de tiempo, como son los procesos estocásticos, haciendo énfasis principal en las cadenas de Markov. Debido a que las variables latentes son conceptos matemáticos motivados por técnicas usadas en la estadística bayesiana, hablaremos también de conceptos fundamentales de este tipo de inferencia, la cual ha sido muy utilizada por los estadísticos en los últimos tiempos. Además se incluye una sección sobre el método Monte Carlo desde su origen hasta los métodos de simulación más modernos que se desarrollaron para aplicarse en la estadística bayesiana los cuales son los métodos de Monte Carlo vía cadenas de Markov y de esta manera, poder plasmar las ideas intuitivas sobre la utilidad de definir variables auxiliares en un modelo probabilístico como un medio para facilitar su construcción y estimación.

1.1. Procesos Estocásticos

Definición 1 (*Proceso Estocástico*)

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, definimos un proceso estocástico como una colección de variables aleatorias $\{X_t : t \in T\}$, indexados por un conjunto no vacío T . Para cada $t \in T$, X_t toma valores en un conjunto E

llamado *espacio de estados*, el conjunto de índices T representará los tiempos en los que se “observan” las variables de la sucesión.

Para los fines de nuestro trabajo tomamos $T = \{0, 1, 2, \dots\}$ y decimos que $\{X_t : t \in T\}$ es un proceso a tiempo discreto. Generalmente tomaremos como espacio de estados E al conjunto de números reales \mathbb{R} o subconjuntos del mismo.

Un proceso estocástico puede considerarse como una función de dos variables $X : T \times \Omega \rightarrow E$ tal que a la pareja (t, ω) se le asocia el estado $X(t, \omega)$, lo cual también se escribe como $X_t(\omega)$. Para cada valor de t en T , el mapeo $\omega \mapsto X_t(\omega)$ es una variable aleatoria, mientras que para cada ω en Ω fijo, la función $t \mapsto X_t(\omega)$ es llamada *trayectoria* o *realización* del proceso.

Si A es un conjunto de estados, el evento $(X_t \in A)$ corresponde a la situación en donde al tiempo t el proceso toma algún valor dentro del conjunto A . En particular, $(X_t = x)$ es el evento en donde al tiempo t el proceso se encuentra en el estado x .

1.1.1. Cadenas de Markov

Asumiremos que cada variable aleatoria X_t en $\{X_t : t \in T\}$ es absolutamente continua¹ y que existe la función de densidad de probabilidades asociada a cada función de distribución que enunciemos.

Definición 2 *Una cadena de Markov es un proceso estocástico a tiempo discreto $\{X_n : n = 0, 1, \dots\}$, con espacio de estados $E \subset \mathbb{R}$ que satisface la propiedad de Markov, esto es, si $\forall n \geq 1$ y $A_0, A_1, A_2, \dots, A_n$ subconjuntos de Borel en \mathbb{R} se cumple que*

$$\mathbb{P}(X_n \in A_n | X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) = \mathbb{P}(X_n \in A_n | X_{n-1} \in A_{n-1}).$$

La distribución de probabilidades $\pi(A) \equiv \mathbb{P}(X_0 \in A)$ se llama *distribución inicial*. En particular si $A_i = \{x_i\}$, $i = 0, 1, \dots, n - 1$.

$$\mathbb{P}(X_n \in A_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_n \in A_n | X_{n-1} = x_{n-1}).$$

¹Decimos que X es absolutamente continua si existe una función $f : \mathbb{R} \rightarrow \mathbb{R}$, $f \geq 0$, tal que $\int_{-\infty}^z f(x)dx = \mathbb{P}(X \leq z)$.

A la familia $\{\mathbb{P}(X_n \in A | X_{n-1} = x); n \in \mathbb{N}, x \in E\}$ se le llama *familia de probabilidades de transición* de la cadena. Describe la evolución de la misma en el tiempo.

Si para cada n y $m \in \{0, 1, \dots\}$ $n \neq m$ tenemos que $\mathbb{P}(X_n \in A | X_{n-1} = x) = \mathbb{P}(X_m \in A | X_{m-1} = x)$, es decir $\mathbb{P}(X_n \in A | X_{n-1} = x)$, no depende de n , decimos que la cadena es homogénea (con respecto al tiempo) y denotamos $P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x)$.

Si $A = (-\infty, x]$, la notación $P(x, y)$ puede ser usada como sigue

$$P(x, y) = \mathbb{P}(X_{n+1} \leq y | X_n = x) = \mathbb{P}(X_1 \leq y | X_0 = x), \forall x, y \in E.$$

Para cada valor de $x \in E$, $P(x, y)$ es una función de distribución de probabilidades. Asumiendo la existencia de una densidad de probabilidades $p(x, y)$ asociada a $P(x, y)$, esta densidad condicional puede obtenerse como

$$p(x, y) = \frac{\partial P(x, y)}{\partial y}, \text{ para } x, y, \in E.$$

Esta densidad puede ser usada para definir las probabilidades de transición de la cadena en lugar de $P(x, A)$.

Cuando se trata de un espacio de estados E discreto, es útil identificar a $P(x, \{y\}) = P(x, y)$. A esta función se le conoce con el nombre de probabilidad de transición y satisface:

- $P(x, y) \geq 0, \forall x, y \in E;$
- $\sum_{y \in E} P(x, y) = 1, \forall x \in E.$

De nuevo en el contexto de cadenas con espacio de estados E , no necesariamente un conjunto numerable, la probabilidad de transición en m pasos se expresa como

$$P^m(x, y) = \mathbb{P}(X_{m+n} \leq y | X_n = x), \text{ para } x, y, \in E.$$

y el kernel (o densidad) de transición en m pasos está dado por

$$p^m(x, y) = \frac{\partial P^m(x, y)}{\partial y}, \text{ para } x, y \in E.$$

Dadas $P^m(z, y)$ y $p^n(x, z)$, se satisface la siguiente relación

$$P^{m+n}(x, y) = \int_{-\infty}^{\infty} P^m(z, y)p^n(x, z)dz, \quad m, n \geq 0, x, y \in E.$$

Esta es la versión de las ecuaciones de *Chapman-Kolmogorov* en espacio de estados continuo. Para $m = 1$ se reduce a

$$P^{n+1}(x, y) = \int_{-\infty}^{\infty} P(z, y)p^n(x, z)dz, \quad n \geq 0, x, y \in E.$$

La distribución marginal $\mathbb{P}(X_n \leq x)$ en el paso n , tiene densidad $\pi^{(n)}$ ($n \geq 0$) que puede ser obtenida de la distribución marginal en el paso anterior con esta expresión

$$\pi^{(n)}(y) = \int_{-\infty}^{\infty} p(x, y)\pi^{(n-1)}(x)dx.$$

Definición 3 *La densidad de probabilidades π correspondiente a la distribución estacionaria o invariante de una cadena de Markov con densidad de transición $p(x, y)$ debe satisfacer*

$$\pi(y) = \int_{-\infty}^{\infty} \pi(x)p(x, y)dx.$$

Esto significa que si $Q_0(x) \equiv \mathbb{P}(X_0 \leq x)$ es una distribución inicial para la cadena y

$$Q_0(x) = \int_{-\infty}^x \pi(z)dz,$$

entonces $\mathbb{P}(X_n \leq x) = \int_{-\infty}^x \pi(z)dz$ o bien la distribución de X_n no cambia con n y tiene por densidad a π . En otras palabras, esta distribución de probabilidad no cambia con el paso del tiempo; de ahí el nombre de estacionaria o invariante.

Para una cadena de Markov con espacio de estados E , un conjunto a lo más numerable, existen resultados que garantizan que si la cadena es positivo-recurrente, entonces existe una distribución invariante $\pi(\cdot)$. Si la cadena es irreducible y aperiódica entonces

$$\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y), \forall x, y \in E,$$

donde para E a lo más numerable $P(x, y) = \mathbb{P}(X_n = y | X_{n-1} = x)$. Entonces, en este caso, independientemente de la distribución inicial de la cadena, $\pi^{(n)}$ se aproxima a π cuando $n \rightarrow \infty$, es decir

$$\begin{aligned} \lim_{n \rightarrow \infty} \pi^n(y) &= \lim_{n \rightarrow \infty} \sum_{x \in E} P^n(x, y) \pi(x) = \sum_{x \in E} \lim_{n \rightarrow \infty} P^n(x, y) \pi(x) \\ &= \sum_{x \in E} \pi(y) \pi(x) = \pi(y), \end{aligned}$$

donde se asume válido el intercambio de límite por suma. En este sentido, a π se le conoce también como distribución límite.

Para una cadena de Markov $\{X_t : t \in T\}$ donde cada X_t es absolutamente continua y con espacio de estados $E \subset \mathbb{R}$, también existen resultados sobre convergencia análogos (véase por ejemplo, Robert y Casella, 1999, Teorema 4.6.5 página 162). No obstante, los conceptos de positivo-recurrencia e irreducibilidad para este caso cambian. Para revisar estos conceptos, pueden verse los libros de Gamerman (1997) página 111 y Robert y Casella (1999) capítulo 4.

1.2. Estadística Bayesiana

1.2.1. Proceso de aprendizaje

La estadística Bayesiana tiene su origen en los estudios del matemático británico (y ministro presbiteriano) Thomas Bayes (1702-1761); él enunció y demostró el resultado que hoy en día conocemos como el “*Teorema de Bayes*”.

En la *estadística clásica o frecuentista* se considera el problema de estimación de un parámetro, o vector de parámetros, θ de una distribución $f(x; \theta)$.

A través de una muestra aleatoria x_1, x_2, \dots, x_n se usan distintos métodos para estimar el valor de θ .

En algunos textos de Estadística Bayesiana como en el libro de J.M. Bernardo (1981), el enfoque Bayesiano de la estadística se aborda como un proceso de aprendizaje, es decir el proceso que permite incorporar al análisis de un problema de decisión la información proporcionada por datos experimentales relacionados con sus sucesos inciertos relevantes.

Este proceso exige precisar, mediante la función de verosimilitud, la relación existente entre los datos y los sucesos inciertos; posteriormente describir mediante una distribución de probabilidad la *información inicial* que se posee sobre las posibilidades de ocurrencia de los sucesos inciertos. La incertidumbre respecto al valor de θ presente en un modelo dado $p(x|\theta)$, es representada a través de una distribución de probabilidad $p(\theta)$ sobre los posibles valores del parámetro desconocido θ (que como ya mencionamos puede ser multidimensional) que define al modelo.

La distribución $p(\theta)$ suele conocerse en la literatura como distribución *a priori* y refleja la información que el observador tiene acerca de θ . El Teorema de Bayes establece que

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

donde la distribución final o *a posteriori* $p(\theta|\mathbf{x})$ describe la información que se posee sobre los sucesos inciertos después de incorporar a la información inicial sobre θ y la información que proporcionan los resultados experimentales contenida en los datos $\mathbf{x} = (x_1, \dots, x_n)$. De esta forma, la distribución *a priori* representa la información de θ antes de tomar la muestra y la distribución *a posteriori* toma en cuenta la información de la muestra.

Este proceso de aprendizaje constituye la base de todo procedimiento estadístico Bayesiano: hacer inferencia sobre el valor (o los valores) de θ se reduce básicamente a determinar su distribución *a posteriori*. Los problemas de *predicción* se resuelven como una extensión natural de los problemas de inferencia mediante el uso del *teorema de probabilidad total*.

En Bernardo (1981), se establece que:

“La reacción natural de cualquiera que tenga que tomar una decisión cuyas consecuencias dependen de la magnitud de una cantidad desconocida θ es intentar reducir su incertidumbre obteniendo más información sobre su valor. El problema central de la inferencia estadística es el de proporcionar una metodología que permita asimilar la información que resulte accesible, con objeto de mejorar nuestro conocimiento del mundo real.”

Por otra parte, del Teorema de Bayes

$$\begin{aligned} p(\theta|\mathbf{x}) &= \frac{p(\theta, \mathbf{x})}{p(\mathbf{x})} \\ &= \frac{p(\mathbf{x}|\theta)p(\theta)}{\int_{\Theta} p(\mathbf{x}|\theta)p(\theta)d\theta} \\ &= \frac{L(\mathbf{x}|\theta)p(\theta)}{\int_{\Theta} L(\mathbf{x}|\theta)p(\theta)d\theta}, \end{aligned}$$

sin pérdida de generalidad, podemos escribir que

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta) = L(\mathbf{x}|\theta)p(\theta) .$$

donde “ \propto ” significa “proporcional a” ó bien “igual salvo por un factor que no depende de θ ”. De manera que si dos modelos probabilísticos son proporcionales para todo θ , corresponden necesariamente a la misma distribución.

En efecto, la constante de proporcionalidad puede determinarse en cualquier momento utilizando la propiedad de que, por ser $p(\theta|\mathbf{x})$ una densidad de probabilidades, tenemos que $\int_{\Theta} p(\theta|\mathbf{x})d\theta = 1$ en el caso continuo y $\sum_{\Theta} p(\theta|\mathbf{x}) = 1$ en el caso discreto. Específicamente, si $p(\theta|\mathbf{x}) = c \times p(\mathbf{x}|\theta)p(\theta)$

$$\int_{\Theta} p(\theta|\mathbf{x}) = c \int_{\Theta} p(\mathbf{x}|\theta)p(\theta)d\theta = 1,$$

luego

$$c = \frac{1}{\int_{\Theta} p(\mathbf{x}|\theta)p(\theta)d\theta} = \frac{1}{p(\mathbf{x})}.$$

Es decir, la cantidad $c = \frac{1}{\int_{\Theta} p(\mathbf{x}|\theta)p(\theta)d\theta}$ solo es una constante con respecto a θ , la cual garantiza que

$$\int_{\Theta} p(\theta|\mathbf{x})d\theta = 1.$$

Si $p(\theta)$ es una constante entonces

$$p(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)$$

y como consecuencia al encontrar el estimador máximo verosímil estaríamos encontrando la moda de la distribución *a posteriori* de θ .

Ejemplo 1 *Inferencia Clásica e Inferencia Bayesiana.*

Sea θ la probabilidad de que caiga águila en el lanzamiento de monedas. Se quiere hacer el contraste

$$H_0 : \theta = 1/2 \quad v.s. \quad H_1 : \theta > 1/2$$

al nivel de significancia $\alpha = .05$.

Supongamos que se tira la moneda y salen 9 águilas y 3 soles. No se conoce la verosimilitud de los datos. Hay dos posibilidades:

(1) Se ha fijado el número de lanzamientos de la moneda en 12. Esto sugiere si X denota el número de águilas obtenidas, $X|\theta \sim \mathcal{B}in(12, \theta)$ y por tanto

$$L(X = 9|\theta) = \binom{12}{9} \theta^9 (1 - \theta)^3.$$

Entonces bajo H_0 , es decir, asumiendo que $\theta = 1/2$ el *p-valor* es

$$p_1 = \sum_{x=9}^{12} \binom{12}{x} (1/2)^x (1/2)^{12-x}$$

$p_1 \approx .073$, por tanto no se rechaza la hipótesis nula.

(2) Se propone que X registre el número de águilas hasta que salga el 3^{er} sol. Luego $X|\theta \sim \mathcal{NB}(3, \theta)$ y

$$L(X = 9|\theta) = \binom{3 + 9 - 1}{9} \theta^9 (1 - \theta)^3.$$

Entonces el p -valor es

$$p_2 = \sum_{x=9}^{\infty} \binom{3 + x - 1}{x} (1/2)^3 (1/2)^x = 1 - \sum_{x=0}^8 \binom{3 + x - 1}{x} (1/2)^3 (1/2)^x$$

$p_2 \approx .0327$, por tanto se rechaza la hipótesis nula.

Como hemos visto, el p -valor fue distinto y la razón es que para hacer un contraste de significancia se tiene que especificar el espacio muestral y éste es diferente en los 2 casos planteados. La inferencia clásica depende del espacio muestral y luego de los datos que pudieramos haber observado.

Desde un punto de vista Bayesiano, podemos usar $p(\theta|\mathbf{x})$ como sigue. Supongamos una distribución *a priori* $\mathcal{U}(0, 1)$ para θ . Lo anterior significa que no se sabe mucho de la moneda; también pudimos haber propuesto a una distribución Beta simétrica $\mathcal{B}e(1/2, 1/2)$ (en tal caso la distribución final $p(\theta|\mathbf{x})$ tendría forma analítica similar). Luego la distribución *a posteriori* es

$$\begin{aligned} p(\theta|\mathbf{x}) &\propto \binom{12}{9} \theta^9 (1 - \theta)^3 \\ &\propto \theta^9 (1 - \theta)^3 \\ &\propto \theta^{10-1} (1 - \theta)^{4-1} \end{aligned}$$

por tanto $\theta|\mathbf{x} \sim \mathcal{B}e(10, 4)$.

Se puede mostrar que $\mathbb{P}(H_1|\mathbf{x}) = \mathbb{P}(\theta > 1/2 | \mathbf{x}) \approx 0.9539$, lo cual no brinda evidencia a favor de H_0 .

También podemos calcular la media de la distribución *a posteriori*. Recordemos que si $X \sim \mathcal{B}e(\alpha, \beta)$, entonces $\mathbb{E}[X] = \alpha/(\alpha + \beta)$; así

$$\mathbb{E}[\theta|\mathbf{x}] = 10/(10 + 4) = 5/7 > 1/2.$$

Nota: Este no es un contraste de hipótesis formal, para mayores detalles sobre contrastes de hipótesis usando inferencia Bayesiana se puede consultar Bernardo y Smith (1994).

Cuando actualizamos nuestro conocimiento acerca de θ mediante la información que aporta una muestra de datos obtenidos, la información total que tenemos acerca de nuestro modelo está contenida en la distribución *a posteriori*

$$p(\theta|z_1) \propto p(\theta)p(z_1|\theta),$$

la cual a su vez se convierte en una nueva distribución *a priori* cuando se cuente con nuevos datos; es decir, si z_2 es un conjunto de nuevos datos que condicionados al valor de θ son independientes de z_1 , entonces

$$p(\theta|z_1, z_2) \propto p(\theta|z_1)p(z_2|\theta),$$

ya que esto es equivalente a haber observado ambos conjuntos de datos desde un principio, pues

$$p(\theta|z_1, z_2) \propto p(\theta)p(z_1, z_2|\theta).$$

Vemos que esto es cierto dado que $p(z_1, z_2|\theta) = p(z_1|\theta)p(z_2|\theta)$ y por tanto

$$\begin{aligned} p(\theta|z_1, z_2) &\propto p(\theta)p(z_1, z_2|\theta) \\ &\propto p(\theta)p(z_1|\theta)p(z_2|\theta) \\ &\propto p(\theta|z_1)p(z_2|\theta). \end{aligned}$$

1.2.2. Predicción

Una de las grandes ventajas de la estadística Bayesiana es que podemos hacer inferencia con tan sólo la información que se tiene de juicios propios o de la experiencia sin presencia de alguna muestra de datos. Si usamos la distribución *a priori*, que representa la información que tenemos acerca del parámetro θ , podemos conocer la probabilidad de que se genere una

muestra \mathbf{x} . Al tener una distribución *a priori* $p(\theta)$ y utilizando el resultado de probabilidad total podemos obtener

$$\begin{aligned} p(\mathbf{x}) &= \int_{\theta \in \Theta} p(\mathbf{x}, \theta) d\theta \\ &= \int_{\theta \in \Theta} p(\mathbf{x}|\theta)p(\theta) d\theta. \end{aligned}$$

A $p(\mathbf{x})$ se le conoce como distribución predictiva *a priori*.

Naturalmente podemos hacer predicción sobre observaciones futuras \tilde{x} del fenómeno aleatorio si contamos con más elementos como una muestra \mathbf{x} , un modelo probabilístico $p(\cdot|\theta)$ y una distribución *a posteriori*. Veamos cómo estos elementos inducen una distribución conjunta para \tilde{x} y θ condicionada a la muestra \mathbf{x} .

$$\begin{aligned} p(\tilde{x}, \theta|\mathbf{x}) &= \frac{p(\tilde{x}, \theta, \mathbf{x})}{p(\mathbf{x})} \\ &= \frac{p(\tilde{x}|\theta, \mathbf{x})p(\theta, \mathbf{x})}{p(\mathbf{x})} \\ &= \frac{p(\tilde{x}|\theta, \mathbf{x})p(\theta|\mathbf{x})p(\mathbf{x})}{p(\mathbf{x})} \\ &= p(\tilde{x}|\theta, \mathbf{x})p(\theta|\mathbf{x}). \end{aligned}$$

Notemos que $p(\tilde{x}|\theta, \mathbf{x}) = p(\tilde{x}|\theta)$ puesto que \mathbf{x} y \tilde{x} son independientes dado el valor de θ .

Obtengamos la distribución marginal de \tilde{x} a partir de la función de distribución conjunta $p(\tilde{x}, \theta|\mathbf{x})$. De la ecuación de arriba tenemos que

$$\begin{aligned} p(\tilde{x}|\mathbf{x}) &= \int_{-\infty}^{\infty} p(\tilde{x}, \theta|\mathbf{x}) d\theta \\ &= \int_{-\infty}^{\infty} p(\tilde{x}|\theta)p(\theta|\mathbf{x}) d\theta. \end{aligned}$$

A la función de distribución $p(\tilde{x}|\mathbf{x})$ se le conoce como distribución predictiva *a posteriori*.

Ejemplo 2 *Se desea predecir el número de águilas en 10 lanzamientos adicionales usando la misma moneda del Ejemplo 1. Por tanto*

$\tilde{x}|\theta \sim \mathcal{B}in(10, \theta)$, y entonces

$$\begin{aligned} p(\tilde{x}|\mathbf{x}) &= \int_{\Theta} p(\tilde{x}|\theta)p(\theta|\mathbf{x})d\theta \\ &= \int_0^1 \binom{10}{\tilde{x}} \theta^{\tilde{x}}(1-\theta)^{10-\tilde{x}} \times \frac{1}{B(10,4)} \theta^{10-1}(1-\theta)^{4-1} d\theta \\ &= \binom{10}{\tilde{x}} \frac{1}{B(10,4)} \int_0^1 \theta^{10+\tilde{x}-1}(1-\theta)^{14-\tilde{x}-1} d\theta \\ &= \binom{10}{\tilde{x}} \frac{B(10+\tilde{x}, 14-\tilde{x})}{B(10,4)}. \end{aligned}$$

la cual es conocida como la distribución *Beta-Binomial*.

Otra cuestión de interés sería poder calcular la media de $\tilde{x}|\mathbf{x}$ y esto se puede hacer sin tener que evaluar la distribución predictiva. Sólo basta recordar que

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$$

para algunas variables X y Y .

Sabemos que $\mathbb{E}[\tilde{x}|\theta] = (10)\theta$ (pues si $X \sim \mathcal{B}in(n, p)$ entonces $\mathbb{E}[X] = np$). Por el Ejemplo 1 sabemos que $\mathbb{E}[\theta|\mathbf{x}] = 5/7$. Luego usando lo anterior

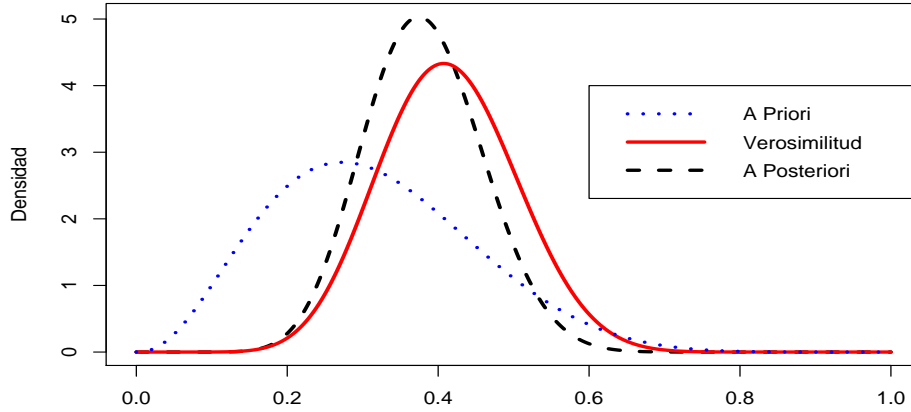


Figura 1.1: La distribución *a priori* es una Beta al igual que la distribución *a posteriori* (sólo que con distintos parámetros) al ser familias conjugadas.

$$\begin{aligned}
 \mathbb{E}[\tilde{X}|\mathbf{x}] &= \mathbb{E}[\mathbb{E}[\tilde{X}|\theta]|\mathbf{x}] \\
 &= \mathbb{E}[(10)\theta|\mathbf{x}] \\
 &= (10)\mathbb{E}[\theta|\mathbf{x}] \\
 &= (10)(5/7) \\
 &= 50/7.
 \end{aligned}$$

Por tanto $\mathbb{E}[\tilde{X}|\mathbf{x}] = 50/7 \approx 7.1428$.

Definición 4 Si \mathcal{F} es una familia paramétrica $\mathcal{F} = \{p(\cdot|\theta) : \theta \in \Theta\}$, y \mathcal{P} es una colección de distribuciones $\mathcal{P} = \{p(\theta) : \theta \in \Theta\}$, decimos que \mathcal{P} es conjugada con respecto a \mathcal{F} si

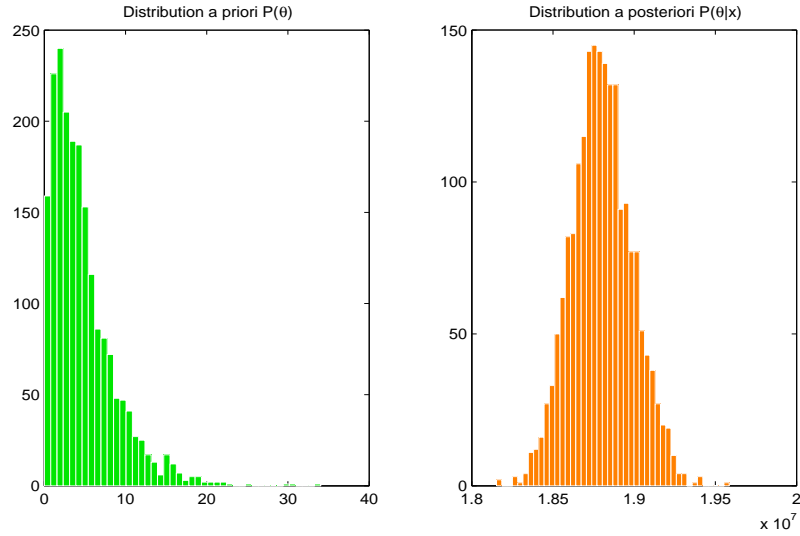


Figura 1.2: Comparación entre la distribución *a priori* y *a posteriori*. Ambas se distribuyen Gamma pero los parámetros cambian al observar la muestra.

$$\forall p(\cdot|\theta) \in \mathcal{F} \text{ y } p(\theta) \in \mathcal{P} \text{ se tiene que } p(\theta|x) \in \mathcal{P}.$$

En la figura 1.1 se muestra una gráfica que ilustra la definición anterior donde se puede apreciar una distribución *a priori* y *a posteriori* pertenecientes a la misma familia (Beta en este caso).

Ejemplo 3 (Modelo Poisson)

Se postula que las variables x_1, \dots, x_n son independientes y condicionalmente a λ , tienen distribución $\mathcal{P}o(x|\lambda)$. Además $\lambda \sim \mathcal{G}a(\alpha, \beta)$. La función de densidad *a posteriori* de λ es,

$$\begin{aligned} p(\lambda | x_1, \dots, x_n) &= \frac{p(x_1, \dots, x_n | \lambda)p(\lambda)}{\int_0^\infty p(x_1, \dots, x_n | \lambda)p(\lambda)d\lambda} \\ &= \frac{\prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^\infty \prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda} \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda^{n\bar{x}+\alpha-1} e^{-(n+\beta)\lambda}}{\int_0^\infty \lambda^{n\bar{x}+\alpha-1} e^{-(n+\beta)\lambda} d\lambda} \\
&= \frac{(n+\beta)^{\sum x_i + \alpha}}{\Gamma(n\bar{x} + \alpha)} \lambda^{\sum x_i + \alpha - 1} e^{-(n+\beta)\lambda}.
\end{aligned}$$

Por lo tanto la densidad *a posteriori* para λ es $\mathcal{G}a(\sum_{i=1}^n x_i + \alpha, n + \beta)$. Esto significa que la distribución Gamma es conjugada con respecto al modelo probabilístico de Poisson; en la figura 1.2 se muestra un histograma de un conjunto de datos con distribución Gamma pero con diferentes parámetros diferenciando la distribución *a priori* de la *a posteriori* como en el ejemplo anterior.

En particular, un caso importante en el que se pueden construir fácilmente familias conjugadas se presenta al utilizar distribuciones pertenecientes a la familia exponencial, es decir familias que se pueden reexpresar como

$$p(x|\theta) = g(\theta)h(x)e^{c(\theta)s(x)}.$$

Denotemos a \mathcal{F} como la familia exponencial y x_1, \dots, x_n una muestra de $p(\mathbf{x}|\theta)$. La verosimilitud es

$$p(\mathbf{x}|\theta) = g(\theta)^n \prod_{i=1}^n h(x_i) e^{c(\theta) \sum_{j=1}^n s(x_j)}.$$

De esta manera, si elegimos una distribución *a priori* de una familia \mathcal{P} de la forma

$$p(\theta) \propto g(\theta)^{n_0} e^{s_0 c(\theta)},$$

entonces la distribución *a posteriori* es de la forma

$$\begin{aligned}
p(\theta|\mathbf{x}) &\propto g(\theta)^{n_0+n} e^{(s_0 + \sum_{j=1}^n s(x_j))c(\theta)} \\
&= g(\theta)^{n_1} e^{s_1 c(\theta)}.
\end{aligned}$$

Por tanto la distribución *a posteriori* también pertenece a \mathcal{P} . De hecho la familia conjugada de una familia exponencial es a su vez una familia exponencial.

Ejemplo 4 (*Familias Conjugadas*)

$$\text{Sea } p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{n}{x} (1-\theta)^n e^{x \log(\frac{\theta}{1-\theta})},$$

que en la notación de la familia exponencial corresponde a

$$h(x) = \binom{n}{x}, \quad g(\theta) = (1-\theta)^n, \quad s(x) = x \quad y \quad c(\theta) = \log\left(\frac{\theta}{1-\theta}\right).$$

Con esta notación, la distribución inicial conjugada toma la forma

$$p(\theta) \propto [(1-\theta)^n]^{n_0} e^{s_0 \log(\frac{\theta}{1-\theta})} \propto \theta^{s_0} (1-\theta)^{nn_0-s_0},$$

que corresponde al kernel de una distribución Beta.

1.2.3. Distribuciones No Informativas

Como mencionamos anteriormente, nuestra distribución *a priori* refleja nuestro conocimiento acerca de θ pero en ocasiones esta información es vaga o sencillamente no se “quiere contaminar” al modelo con la información subjetiva que se pueda aportar. Esto ha introducido el concepto de distribuciones *a priori* llamadas *no informativas*. En algunos textos se menciona el *principio de la razón insuficiente* el cual establece que si no hay información para diferenciar entre valores diferentes de θ , se debe de dar la misma probabilidad a todos los valores. Esto sugiere una distribución *a priori* uniforme para θ

$$p(\theta) \propto 1.$$

Si el soporte de θ es infinito la distribución *a priori* será *impropia*, es decir

$$\int_{\Theta} p(\theta) d\theta = \infty$$

Buscando obtener invarianza ante transformaciones de los parámetros, Jeffreys (1961) propuso usar como distribución inicial no informativa a

$$p(\theta) \propto |I(\theta)|^{1/2},$$

donde

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \log L(\mathbf{x}|\theta) \right]$$

es conocida como *la información de Fisher*. Aunque ésta es una distribución no informativa utilizada en el caso unidimensional, Bernardo y Smith (1994) señalan que puede no ser muy útil en el caso multivariado. El método de Berger y Bernardo (1992)² es otra solución muy usada.

El lector interesado en el enfoque Bayesiano de la Estadística puede consultar los libros de Bernardo (1981), Bernardo y Smith (1994) y los libros citados en la parte de la bibliografía referentes a este tema.

1.3. Variables latentes

El método de variables latentes consiste en notar que en ocasiones es posible simular más fácil y eficientemente de una distribución $f(\mathbf{x}, \lambda|\Theta)$ que de la distribución $f(\mathbf{x}|\Theta)$, donde el nuevo “parámetro” λ puede ser esencialmente cualquier variable aleatoria; a esta variable λ se le llama *variable latente*. Es evidente que al generar una muestra de la distribución conjunta $f(\mathbf{x}, \lambda|\Theta)$ automáticamente obtenemos una muestra de la distribución marginal $f(\mathbf{x}|\Theta)$ que nos interesaba originalmente.

Ejemplo 5 (*Aplicación de las variables latentes*)

Supongamos que deseamos simular un vector aleatorio $\mathbf{x} \sim \mathcal{N}b(\alpha, p)$. Esta simulación puede realizarse usando el hecho de que si $\lambda \sim \mathcal{G}a(\alpha, p/(1-p))$ y $\mathbf{x}|\lambda \sim \mathcal{P}o(\lambda)$, entonces $\mathbf{x} \sim \mathcal{N}b(\alpha, p)$, ya que es más sencillo simular vectores aleatorios que se distribuyan *Gamma* o *Poisson*. En este ejemplo, una de las posibles variables latentes sería λ ya que

$$\int f(\mathbf{x}|\lambda) \times f(\lambda) d\lambda = \int f(\mathbf{x}, \lambda) d\lambda = f(\mathbf{x}).$$

² “Distribuciones de Referencia”

Verifiquemos que esto es cierto; hagamos por el momento el cambio de variable $\beta = p/(1 - p)$. Entonces

$$\begin{aligned}
f(\mathbf{x}) &= \int f(\mathbf{x}|\lambda) \times f(\lambda)d\lambda \\
&= \int_0^\infty \left(\frac{\lambda^x e^{-\lambda}}{x!} \right) \times \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \right) d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x!} \int_0^\infty \lambda^{x+\alpha-1} e^{-\lambda\beta-\lambda} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x!} \frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}} \times \int_0^\infty \frac{(\beta+1)^{x+\alpha}}{\Gamma(x+\alpha)} \lambda^{x+\alpha-1} e^{-\lambda(\beta+1)} d\lambda \\
&= \frac{\beta^\alpha}{(\beta+1)^{x+\alpha}} \frac{\Gamma(x+\alpha)}{\Gamma(\alpha) x!} \\
&= \left(\frac{\beta}{\beta+1} \right)^\alpha \left(\frac{1}{\beta+1} \right)^x \frac{(x+\alpha-1)!}{(\alpha-1)! x!} \\
&= \binom{\alpha+x-1}{\alpha-1} \left(\frac{1}{\beta+1} \right)^x \left(\frac{\beta}{\beta+1} \right)^\alpha \tag{1.1}
\end{aligned}$$

Por lo tanto $\mathbf{x} \sim \mathcal{Nb}(\alpha, p)$, $p = \frac{\beta}{\beta+1}$. Si se conocieran las distribuciones condicionales completas, es decir $f(x|\lambda)$ y $f(\lambda|x)$, se puede realizar la simulación usando el algoritmo Gibbs Sampler. No obstante, en los párrafos que anteceden observamos que a partir de variables que se distribuyen Gamma y Poisson podemos generar variables que se distribuyen Binomial Negativa; cabe mencionar que la distribución Binomial Negativa es un caso particular de la distribución *Poisson-Gamma*, usada frecuentemente en la estadística Bayesiana y como alternativa a la distribución Poisson cuando se requiere modelar datos discretos con sobredispersión (es decir, datos para los cuales la varianza muestral excede a la varianza del modelo).

1.4. Algoritmo EM

El algoritmo EM (*Expectation-Maximization*) es un procedimiento iterativo para calcular la moda de la distribución *a posteriori* o el estimador máximo verosímil para datos incompletos, es decir, cuando sólo un subconjunto de los datos está disponible o cuando hay datos que no pueden ser observados. Dempster, Laird y Rubin (1977) demostraron que la aplicación del algoritmo EM es muy extensa y son en gran parte responsables de la popularización de este método en estadística.

En la formulación usual del algoritmo EM, el vector “completo” de datos \mathbf{x} es conformado por “datos observados” \mathbf{y} (algunas veces llamados datos incompletos) y por datos “no observables” \mathbf{w} , es decir que los elementos de \mathbf{w} se pueden considerar como variables latentes. El algoritmo EM provee un procedimiento iterativo para calcular los estimadores de máxima verosimilitud basado sólo en los datos observados. Cada iteración del algoritmo EM consiste de dos pasos: el paso **E** (paso de la esperanza) y el paso **M** (paso de la maximización). Supóngase que $\theta^{(k)}$ denota el valor estimado después de k iteraciones del algoritmo para la moda de la distribución *a posteriori* de nuestro interés $f(\theta|\mathbf{y})$ y sea $f(\theta|\mathbf{y}, \mathbf{w})$ la distribución *a posteriori* aumentada. Por último definamos a $f(\mathbf{w}|\mathbf{y}, \theta^{(k)})$ como la distribución predictiva condicional de los datos latentes \mathbf{w} dado el valor k -ésimo de la moda *a posteriori* y dados los datos observados, entonces los pasos E y M en la $(k + 1)$ -ésima iteración son

paso – E. Calcular $Q(\theta|\theta^{(k)}) = \mathbb{E}[\log f(\theta|\mathbf{y}, \mathbf{w})]$ con respecto a $f(\mathbf{w}|\mathbf{y}, \theta^{(k)})$,

es decir,

$$Q(\theta|\theta^{(k)}) = \int_{\mathcal{W}} \log f(\theta|\mathbf{y}, \mathbf{w}) f(\mathbf{w}|\mathbf{y}, \theta^{(k)}) d\mathbf{w}$$

y

paso – M. Maximizar $Q(\theta, \theta^{(k)})$ con respecto a θ . Se toma $\theta^{(k+1)}$ como el valor donde $Q(\cdot, \theta^{(k)})$ se maximiza.

Después de ésto regresamos al paso-E. El algoritmo se itera hasta que

$$|Q(\theta^{(k+1)}, \theta^{(k)}) - Q(\theta^{(k)}, \theta^{(k)})| \text{ o bien } \|\theta^{(k)} - \theta^{(k+1)}\|$$

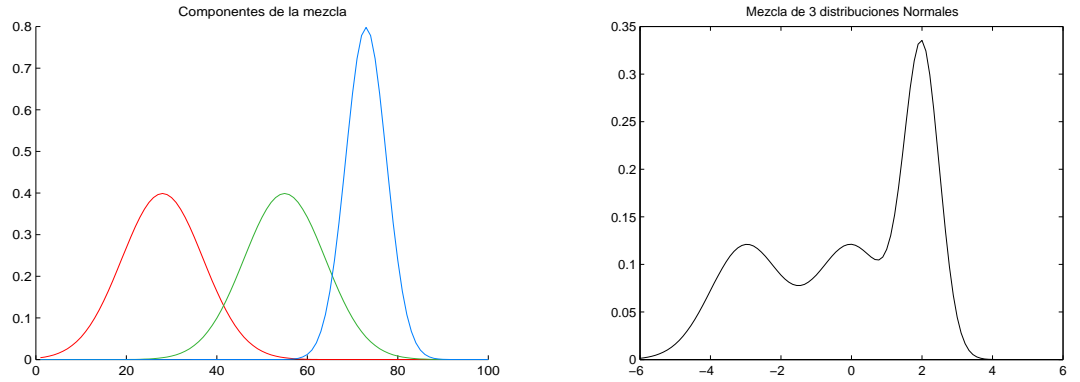


Figura 1.3: Mezcla de 3 distribuciones normales y sus componentes

sean suficientemente pequeño.

Se calcula el valor $\mathbb{E}[\log f(\theta|\mathbf{y}, \mathbf{w})]$ debido a que se desconocen los valores de \mathbf{w} y se considera entonces el valor esperado de esta variable. Al maximizar la función $Q(\theta|\theta^{(k)})$ en la k -ésima iteración del algoritmo, se obtiene un valor $\theta^{(k+1)}$ tal que $f(\theta^{(k+1)}|Y) \geq f(\theta^{(k)}|Y)$ (véase Tanner, M.A. , 1996, página 70).

Ejemplo 6 (Uso del algoritmo EM)

Vamos a mostrar una de las múltiples aplicaciones del algoritmo EM para estimar parámetros en las llamadas *mezclas finitas de distribuciones*. Una mezcla finita de distribuciones es una combinación ponderada de distribuciones de probabilidad

$$f(x|\Theta) = \sum_{j=1}^M \pi_j f_j(x|\theta_j), \quad \Theta = \{\pi_j, \theta_j\}, \quad \sum_j \pi_j = 1.$$

En la figura 1.3 se puede apreciar una mezcla de normales con diferentes medias (lado derecho) y sus componentes individuales (izquierdo). Si quisiéramos estimar estos parámetros mediante máxima verosimilitud vemos que

$$\log L(\Theta|\mathbf{x}) = \log \prod_{i=1}^N \left(\sum_{j=1}^M \pi_j f_j(x_i|\theta_j) \right) = \sum_{i=1}^N \log \left(\sum_{j=1}^M \pi_j f_j(x_i|\theta_j) \right).$$

Es usual que la ecuación $\frac{\partial}{\partial \Theta} \log L(\Theta|x) = 0$, no se pueda resolver analíticamente. Introduzcamos entonces una variable latente y_i asociada a los valores conocidos de la muestra x_i tales que $f(y_i = j|x_i, \Theta)$ representa la probabilidad de que x_i pertenezca a la componente j de la mezcla.

Evidentemente $f(x_i|y_i = j, \Theta) = f_j(x_i|\theta_j)$ y $f(y_i = j|\Theta) = \pi_j$. Así, en esta nueva formulación tendríamos

$$f(x_i|\Theta) = \sum_{j=1}^M f(x_i, y_i = j|\Theta) = \sum_{j=1}^M f(x_i|y_i = j, \Theta) f(y_i = j|\Theta) = \sum_{j=1}^M \pi_j f_j(x_i|\theta_j).$$

En el paso-E calculamos $Q(\Theta, \Theta^{(i)}) = E[\log L(\Theta|x_i, y_i)|\mathbf{x}, \Theta^{(i)}]$

En el paso Paso-M encontramos $\Theta^{(i+1)} = \arg \max_{\Theta} Q(\Theta, \Theta^{(i)})$.

Finalmente iteramos este algoritmo hasta obtener la convergencia deseada para $\Theta^{(i)}$.

Ejemplo 7 (*Distribución Poisson-Gamma*)

En la introducción a la sección 1.3 se habló del uso de las variables latentes para simular números aleatorios de la distribución Binomial Negativa. Una generalización de ésta es la distribución Poisson-Gamma con parámetros $\alpha > 0$, $\beta > 0$ y $\nu > 0$, con densidad dada por

$$f(x|\alpha, \beta, \nu) = \binom{\alpha + x - 1}{\alpha} \left(\frac{\nu}{\beta + \nu} \right)^x \left(\frac{\beta}{\beta + \nu} \right)^{\alpha}; \quad x = 0, 1, \dots$$

Vamos a usar el algoritmo EM para encontrar estimadores máximo verosimiles de α , β y ν dada una muestra y_1, \dots, y_n . Para lo anterior notemos que

$$f(y|\alpha, \beta, \nu) = \int_0^{\infty} f(y|w\nu) \times f(w|\alpha, \beta) dw \quad (1.2)$$

donde $f(y|w\nu) = \frac{(w\nu)^y}{y!}e^{-w\nu}$ y $f(w|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}w^{\alpha-1}e^{-\beta w}$.

Es decir, para una variable latente w con distribución $\text{Gamma}(w|\alpha, \beta)$ la distribución de y condicional a w y ν es $\text{Poisson}(y|w\nu)$. La relación (1.2) se prueba en forma análoga a como se verificó (1.1). Dada una muestra aleatoria $\mathbf{Y} = (y_1, \dots, y_n)$ de datos provenientes de $f(y|\alpha, \beta, \nu)$, consideremos que para cada y_i tenemos una variable latente w_i , de forma que la densidad conjunta de y_i y w_i es

$$\begin{aligned} f(y_i, w_i|\alpha, \beta, \nu) &= p(y_i|w_i, \nu) \times p(w_i|\alpha, \beta) \\ &= f(y_i|w_i, \nu) \times f(w_i|\alpha, \beta) \\ &= \left\{ \frac{(w_i\nu)^{y_i}}{y_i!} e^{-w_i\nu} \right\} \times \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} w_i^{\alpha-1} e^{-\beta w_i} \right\}; \end{aligned}$$

la densidad conjunta (verosimilitud) de los datos completos $\mathbf{X} = (\mathbf{Y}, \mathbf{W}) = ((y_1, w_1), (y_2, w_2), \dots, (y_n, w_n))$ es

$$\begin{aligned} f(\mathbf{X}|\alpha, \beta, \nu) &= \prod_{i=1}^n f(y_i, w_i|\alpha, \beta, \nu) = \prod_{i=1}^n \left\{ f(y_i|w_i, \nu) \times f(w_i|\alpha, \beta) \right\} \\ &= \prod_{i=1}^n \left\{ \left\{ \frac{\nu^{y_i}}{y_i!} \right\} \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} \right\} w_i^{\alpha+y_i-1} e^{-(\nu+\beta)w_i} \right\}. \end{aligned} \quad (1.3)$$

Usando la relación (1.2) encontramos que la densidad conjunta (verosimilitud) de los datos observados \mathbf{Y} es

$$f(\mathbf{Y}|\alpha, \beta, \nu) = \prod_{i=1}^n \left[\left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} \right\} \times \frac{\nu^{y_i}}{y_i!} \times \frac{\Gamma(y_i + \alpha)}{(\beta + \nu)^{y_i + \alpha}} \right]. \quad (1.4)$$

Comenzamos entonces a identificar los ingredientes necesarios para instrumentar el algoritmo EM

(a) $\log\{f(\theta|\mathbf{X})\}$:

Asumiendo una distribución inicial uniforme para α, β, ν ($p(\alpha, \beta, \nu) \propto$

1), de (1.3) obtenemos que el logaritmo de la distribución posterior aumentada $f(\alpha, \beta, \nu | \mathbf{X})$ es

$$\begin{aligned}
 \log f(\alpha, \beta, \nu | \mathbf{X}) &= \sum_{i=1}^n \left[y_i \times \log(\nu) - \log\{y_i!\} \right] \\
 &+ n \left\{ \log(\beta) \times \alpha - \log \Gamma(\alpha) \right\} \\
 &+ \sum_{i=1}^n (\alpha + y_i - 1) \log(w_i) \\
 &- \sum_{i=1}^n (\nu + \beta) w_i
 \end{aligned} \tag{1.5}$$

(b) $f(\mathbf{W} | \mathbf{Y}, \theta^{(k)})$

Usando (1.3) y (1.4), obtenemos que la densidad predictiva condicional de los datos latentes es

$$f(\mathbf{W} | \mathbf{Y}, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}) = \prod_{i=1}^n \left\{ \frac{(\beta^{(k)} + \nu^{(k)})^{y_i + \alpha^{(k)}}}{\Gamma(y_i + \alpha^{(k)})} w_i^{\alpha^{(k)} + y_i - 1} e^{-(\nu^{(k)} + \beta^{(k)}) w_i} \right\}$$

Usaremos esta densidad para integrar (1.5) con respecto a los datos latentes \mathbf{W}

$$\begin{aligned}
Q(\alpha, \beta, \nu | \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}) &= \sum_{i=1}^n \left[y_i \log(\nu) - \log\{y_i!\} \right] \\
&+ n \left[\log(\beta)\alpha - \log(\Gamma(\alpha)) \right] \\
&+ \sum_{i=1}^n (\alpha + y_i - 1) \mathbb{E} \left[\log(w_i) | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)} \right] \\
&- \sum_{i=1}^n (\nu + \beta) \mathbb{E} \left[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)} \right];
\end{aligned}$$

lo anterior describe el paso E del algoritmo. Para hacer el paso M, hay que maximizar $Q(\alpha, \beta, \nu | \alpha^{(k)}, \beta^{(k)}, \nu^{(k)})$ con respecto a α, β, ν .

$$\begin{aligned}
\frac{\partial Q(\alpha, \beta, \nu)}{\partial \nu} &= \frac{1}{\nu} \sum_{i=1}^n y_i - \sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] \\
\text{si y solo si } \frac{1}{\nu^{(k)}} &= \frac{\sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}]}{\sum_{i=1}^n y_i} \\
\text{si y solo si } \nu^{(k+1)} &= \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}]} \quad (1.6)
\end{aligned}$$

$$\frac{\partial Q(\alpha, \beta, \nu)}{\partial \alpha} = n \log(\beta) - \psi(\alpha) + \sum_{i=1}^n \mathbb{E}[\log(w_i) | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] \quad (1.7)$$

$$\frac{\partial Q(\alpha, \beta, \nu)}{\partial \beta} = n \left(\frac{\alpha}{\beta} \right) - \sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] \quad (1.8)$$

De (1.8)

$$\alpha^{(k+1)} = \beta^{(k+1)} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}]$$

ó

$$\beta^{(k+1)} = \frac{\alpha^{(k+1)}}{\frac{1}{n} \sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}]} \quad (1.9)$$

Para resolver $\alpha^{(k+1)}$ numéricamente se tiene que

$$\begin{aligned} \psi(\alpha^{(k+1)}) &= n \log \left[\frac{\alpha^{(k+1)}}{\frac{1}{n} \sum_{i=1}^n \mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}]} \right] \\ &+ \sum_{i=1}^n \mathbb{E}[\log(w_i) | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] \end{aligned} \quad (1.10)$$

$$f(\mathbf{W} | \mathbf{Y}, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}) = \prod_{i=1}^n \left\{ \frac{(\beta^{(k)} + \nu^{(k)})^{y_i + \alpha^{(k)}}}{\Gamma(y_i + \alpha^{(k)})} w_i^{\alpha^{(k)} + y_i - 1} e^{-(\nu^{(k)} + \beta^{(k)})w_i} \right\};$$

por tanto $\{w_i\}_i$ son v.a.'s independientes tales que $w_i \sim Ga(w_i | y_i + \alpha^{(k)}, \beta^{(k)} + \nu^{(k)})$ y entonces sabemos que

$$\mathbb{E}[w_i | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] = \frac{y_i + \alpha^{(k)}}{\beta^{(k)} + \nu^{(k)}}; \quad (1.11)$$

por otro lado, para $w \sim Ga(w | \alpha, \beta)$ tenemos que

$$\mathbb{E}[\log(w)] = \psi(\alpha) - \log(\beta).$$

Así que

$$\mathbb{E}[\log(W_i) | y_i, \alpha^{(k)}, \beta^{(k)}, \nu^{(k)}] = \psi(y_i + \alpha^{(k)}) - \log(\beta^{(k)} + \nu^{(k)}). \quad (1.12)$$

Dados (1.11) y (1.12) podemos evaluar $\alpha^{(k+1)}$, $\beta^{(k+1)}$ y $\nu^{(k+1)}$ recursivamente usando (1.6), (1.9) y (1.10).

Capítulo 2

Series de Tiempo

En este capítulo revisaremos los fundamentos matemáticos más esenciales y necesarios para entender los modelos de series de tiempo. Estos conceptos van desde la estructura probabilística que deben de poseer el conjuntos de datos que deseamos describir mediante este tipo de modelos, la clasificación y características principales de los modelos más utilizados en la práctica, los métodos de diagnóstico, estimación de parámetros y las pruebas estadísticas que nos ayudarán a elegir el modelo más apropiado dentro de un conjunto de modelos posibles. Profundizaremos un poco en las definiciones de estacionaridad ya que este concepto es vital para la construcción de series de tiempo con variables latentes garantizando que estas sigan siendo estacionarias.

Los modelos de series de tiempo proporcionan un método sofisticado para describir y predecir series históricas ya que se basan en la idea de que el conjunto de datos en estudio ha sido generado por un proceso estocástico, con una estructura que puede caracterizarse y describirse. La descripción no se da en función de una relación causa y efecto (como sería en el caso de un modelo de regresión) sino en función de cómo está incorporada la aleatoriedad en el proceso. Otra de las diferencias primordiales entre las series de tiempo y los modelos de regresión es el supuesto de independencia ya que ésta es una hipótesis que no se cumple en las series de tiempo pues en estos datos se presenta autocorrelación (total o parcial) a través del tiempo.

Hablando en términos matemáticos, una serie de tiempo puede entenderse como una realización del proceso estocástico $\{X_t : t \in T\}$, es decir, para $w \in \Omega$

$$x_{t_1} = x_{t_1}(w), x_{t_2} = x_{t_2}(w), \dots$$

y las observaciones $\{x_t : t \in T\}$ son la serie de tiempo.

El conjunto T representa tiempos t_1, t_2, \dots en donde se llevó a cabo la medición u observación. Así entonces, para nosotros T será un conjunto a lo más numerable, es decir, $T = \{t_1, t_2, \dots\}$ y, a menos que se especifique otra cosa, estaremos considerando $T = \mathbb{Z}$ ó $T = \{0, 1, 2, \dots\}$ para nuestro estudio.

Denotemos

$$F_t(a) = \mathbb{P}(X_t \leq a).$$

Supóngase que, para cada $t \in T$, tanto $\mathbb{E}(X_t)$ como $\text{Var}(X_t)$ existen y son finitos; entonces,

$$\mathbb{E}(X_t) = \int_{-\infty}^{\infty} x f(x) dx \equiv \mu_t \quad y$$

$$\text{Var}(X_t) = \int_{-\infty}^{\infty} (x - \mu_t)^2 f(x) dx \equiv \sigma_t^2.$$

Una condición equivalente sería que $\mathbb{E}(X_t^2)$ exista y sea finita para cada $t \in T$. Para verificar esta afirmación recordemos que si $\mathbb{E}(X_t^2)$ es finita, entonces $\mathbb{E}(X_t)$ es finito. En efecto,

$$\infty > \int_{-\infty}^{\infty} x^2 f_{X_t}(x) dx \geq \int_{|x|>1} x^2 f_{X_t}(x) dx > \int_{|x|>1} x f_{X_t}(x) dx$$

y como $\int_{|x|\leq 1} x f_{X_t}(x) dx \leq \int_{|x|\leq 1} f_{X_t}(x) dx < \infty$ entonces $\mathbb{E}(X_t)$ existe y es finita.

Por otra parte

$$\text{Var}(X_t) = \mathbb{E}(X_t^2) - (\mathbb{E}(X_t))^2,$$

entonces $\text{Var}(X_t)$ y $\mathbb{E}(X_t)$ son finitos. Inversamente si $\text{Var}(X_t)$ y $\mathbb{E}(X_t)$ son finitos entonces como $\mathbb{E}(X_t^2) = \text{Var}(X_t) + (\mathbb{E}(X_t))^2$ tenemos que $\mathbb{E}(X_t^2)$ es finito. En resumen $\text{Var}(X_t)$ y $\mathbb{E}(X_t)$ existen y son finitas si y sólo si $\mathbb{E}(X_t^2)$ existe y es finita.

Decimos que el proceso $\{X_t\}_t$ es un proceso de 2º orden si $\forall t \in T$ $\mathbb{E}(X_t^2) < \infty$.

Sea $\{X_t\}_{t \in T}$ un proceso estocástico y para cada $n \in \{1, 2, \dots\}$ consideremos una colección de tiempos $t_1, t_2, \dots, t_n \in T$. Las distribuciones de probabilidad conjuntas del vector $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ están dadas por

$$F_{t_1, t_2, \dots, t_n}(a_1, a_2, \dots, a_n) = \mathbb{P}(X_{t_1} \leq a_1; \dots; X_{t_n} \leq a_n). \quad (2.1)$$

Estas distribuciones se llaman *distribuciones de dimensión finita* del proceso $\{X_t\}_t$.

2.1. Procesos Estacionarios

Definición 5 (*Proceso estacionario*)

Un proceso $\{X_t\}_{t \in T}$ se llama *completamente estacionario* si para todo $n \geq 1$, para cualesquiera t_1, t_2, \dots, t_n elementos de T y para todo τ tal que $t_1 + \tau, t_2 + \tau, \dots, t_n + \tau \in T$, la función de distribución conjunta del vector $(X_{t_1}, \dots, X_{t_n})$ es igual a la distribución conjunta del vector $(X_{t_1 + \tau}, X_{t_2 + \tau}, \dots, X_{t_n + \tau})$

$$\begin{aligned} F_{t_1, \dots, t_n}(a_1, \dots, a_n) &= \mathbb{P}(X_{t_1} \leq a_1; \dots; X_{t_n} \leq a_n) \\ &= \mathbb{P}(X_{t_1 + \tau} \leq a_1; \dots; X_{t_n + \tau} \leq a_n) \\ &= F_{t_1 + \tau, \dots, t_n + \tau}(a_1, \dots, a_n). \end{aligned}$$

En otras palabras, toda la estructura probabilística de un proceso completamente estacionario no cambia al correr en el tiempo (véase la figura 2.1).

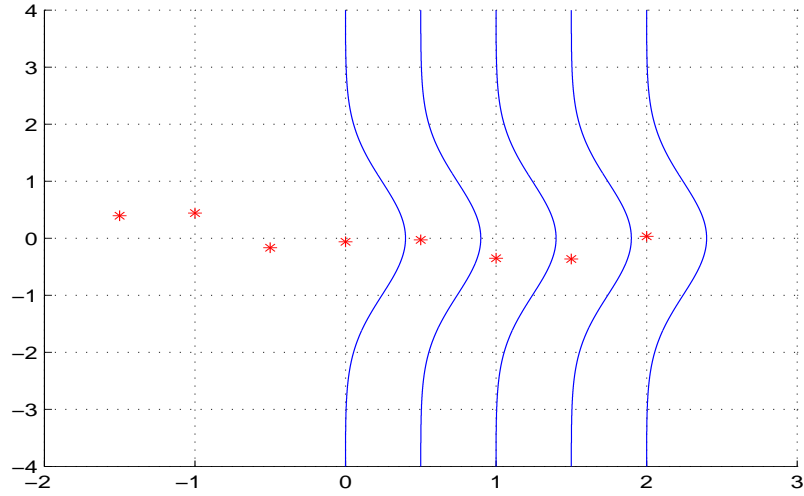


Figura 2.1: Proceso completamente estacionario

Sin embargo para varios fenómenos en la naturaleza tales modelos resultan restrictivos; para modelar datos reales necesitamos “relajar” la definición.

Definición 6 (*Proceso estacionario de 2º orden*)

Un proceso de 2º orden $\{X_t\}_{t \in T}$ se llama débilmente estacionario o estacionario de 2º orden si:

Para cada $n \in \{1, 2, \dots\}$, para cualesquiera $t_1, t_2, \dots, t_n \in T$ y para todo τ tal que $t_1 + \tau, \dots, t_n + \tau \in T$, todos los momentos¹ de orden 1 y 2 del vector $(X_{t_1}, \dots, X_{t_n})$ son iguales a los correspondientes momentos de orden 1 y 2 del vector $(X_{t_1 + \tau}, X_{t_2 + \tau}, \dots, X_{t_n + \tau})$,

$$\mathbb{E}[X_{t_1}^{r_1} \cdots X_{t_n}^{r_n}] = \mathbb{E}[X_{t_1 + \tau}^{r_1} \cdots X_{t_n + \tau}^{r_n}],$$

donde $r_i \in \{0, 1, 2\}$ y $\sum_{i=1}^n r_i \leq 2$.

¹Notemos que como $\{X_t\}_t$ es de segundo orden, estos momentos conjuntos existen y son finitos; esto es consecuencia de la desigualdad de Cauchy-Shwarz.

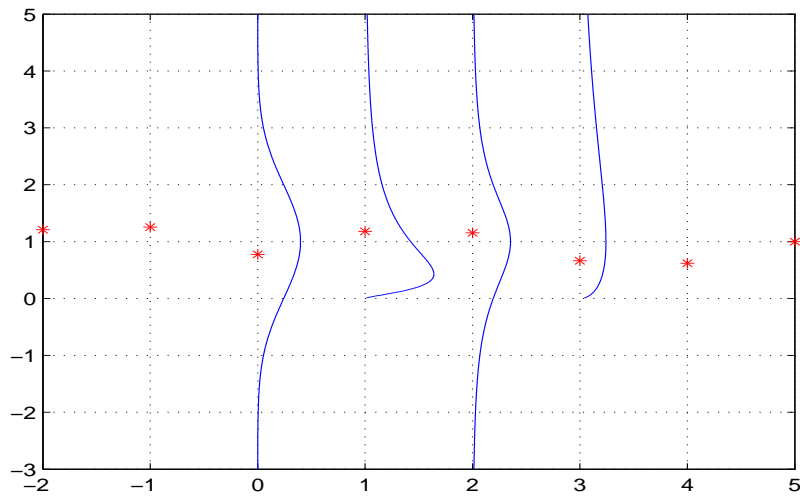


Figura 2.2: Proceso débilmente estacionario

En la figura 2.2 podemos ver un ejemplo de una serie débilmente estacionaria ya que como se puede apreciar, la distribución no es la misma para cada observación sin embargo la media parece ser la misma para todas las observaciones. Por otro lado, en la figura 2.3 se muestra un ejemplo de una serie no estacionaria pues ni la distribución ni la media es la misma para cada observación.

En general se tendrá que:

(i) Si $\{X_t\}_t$ es estacionario fuerte y además es un proceso de 2^o orden, entonces $\{X_t\}_t$ es estacionario de 2^o orden, o bien estacionariedad fuerte y existencia de momentos de 2^o orden implica estacionariedad débil.

(ii) Si $\{X_t\}_t$ es estacionario de 2^o orden, entonces no necesariamente $\{X_t\}_t$ es fuertemente estacionario.

Sea $\{X_t\}_t$ un proceso estacionario de 2^o orden. Entonces, tomando $\tau = -s$

$$\begin{aligned} \text{Cov}[X_t, X_s] &= \text{Cov}[X_{t+\tau}, X_{s+\tau}] \\ &= \text{Cov}[X_{t-s}, X_0], \end{aligned}$$

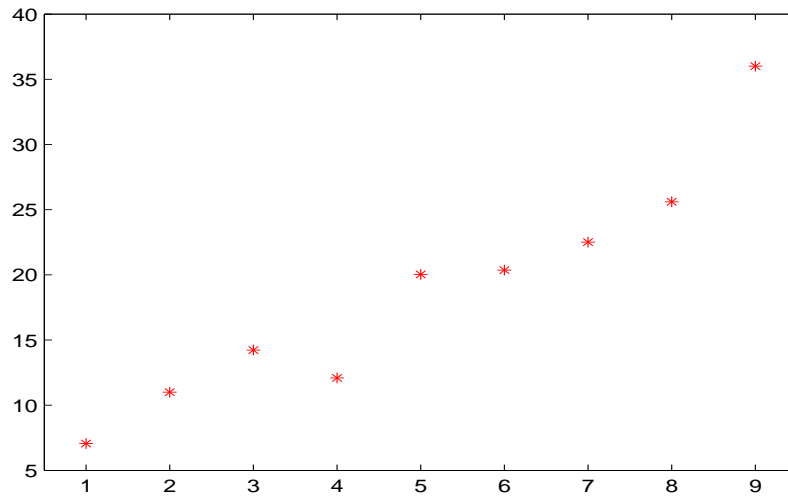


Figura 2.3: Proceso no estacionario

si ahora tomamos $\tau = -t$

$$\begin{aligned} \text{Cov}[X_t, X_s] &= \text{Cov}[X_{t+\tau}, X_{s+\tau}] \\ &= \text{Cov}[X_0, X_{s-t}]. \end{aligned} \quad (2.2)$$

Es decir

$$\begin{aligned} \text{Cov}[X_t, X_s] &= \text{Cov}[X_0, X_{s-t}] \\ &= \text{Cov}[X_{t-s}, X_0]. \end{aligned} \quad (2.3)$$

y estas dos cantidades dependen de s y t sólo a través de $|s - t|$. En otras palabras, para un proceso estacionario de 2º orden la covarianza entre X_s y X_t es función de $|t - s|$. Así por ejemplo, la covarianza entre X_1 y X_5 vale lo mismo que la covarianza entre X_3 y X_7 .

De lo anterior definimos la sucesión de autocovarianza de $\{X_t : t \in T\}$ como

$$S_\tau = \text{Cov}[X_s, X_{s+\tau}] \quad \forall s \in T, \text{ y } \tau \in T \text{ tal que } s + \tau \in T.$$

Notemos que en virtud de las ecuaciones (2.2) y (2.3) esta sucesión es una función par de τ , es decir,

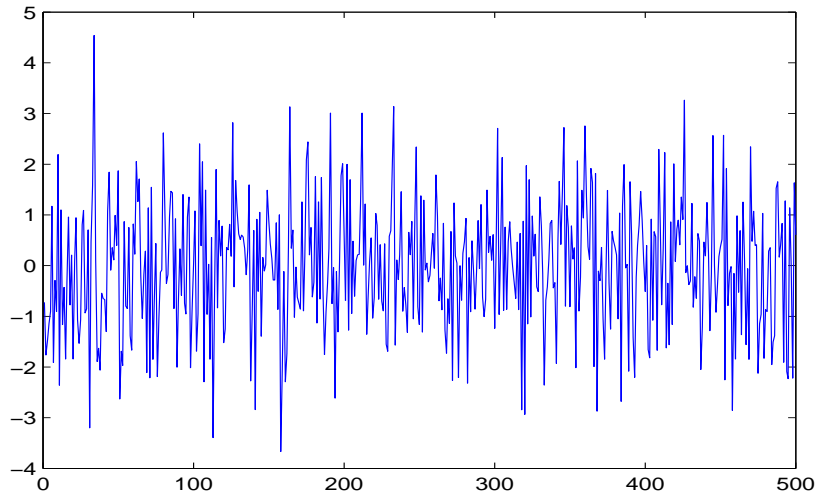


Figura 2.4: Ejemplo de un proceso de ruido blanco

$$S_\tau = S_{-\tau}.$$

Así mismo, se define la sucesión de autocorrelación del proceso $\{X_t\}_t$ como $\rho_\tau = \frac{S_\tau}{S_0}$; $\tau = 0, \pm 1, \pm 2, \dots$

Ahora, veremos algunos ejemplos de procesos estacionarios de 2º orden con los cuales vamos a trabajar.

2.1.1. Ruido Blanco

Sea $\{X_t\}_t$ una sucesión de v.a. no-correlacionadas, tal que $\mathbb{E}[X_t] = \mu$ y $\text{Var}[X_t] = \sigma^2$, $\forall t$, entonces $\text{Cov}[X_t, X_{t+\tau}] = 0 \quad \forall t \in T$ y $\tau \neq 0$, luego

$$S_\tau = \begin{cases} \sigma^2, & \text{si } \tau = 0, \\ 0, & \text{si } \tau \neq 0, \end{cases} \quad \rho_\tau = \begin{cases} 1, & \text{si } \tau = 0, \\ 0, & \text{si } \tau \neq 0. \end{cases}$$

La figura 2.4 muestra un ejemplo de una realización de un proceso de ruido blanco.

2.1.2. Promedios Móviles

El proceso $\{X_t\}_t$ se llama de promedios móviles con orden q (MA(q)) si se puede escribir como

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},$$

donde μ y θ_j son constantes, $j = 1, 2, \dots, q$. El proceso $\{\varepsilon_t\}$ es ruido blanco con $\mathbb{E}(\varepsilon_t) = 0, \forall t$ y $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

Tenemos que $\mathbb{E}(X_t) = \mu$ y el proceso $Y_t = X_t - \mu$ es tal que

$$\begin{aligned} S_\tau &= \mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)] \\ &= \mathbb{E}[Y_t \cdot Y_{t+\tau}]. \end{aligned}$$

Puede verse en Brockwell y Davis (1987) que S_τ es de la forma

$$S_\tau = \begin{cases} \sigma^2 \sum_{j=0}^q \theta_j \theta_{j+|\tau|}, & \text{si } |\tau| \leq q, \\ 0, & \text{si } |\tau| > q. \end{cases}$$

La figura 2.5 nos muestra un ejemplo de una realización de un proceso MA(1).

2.1.3. Autorregresivo

El proceso $\{X_t\}_t$ se llama autorregresivo de orden p si se puede escribir como

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad (2.4)$$

donde μ y $\phi_j, j = 1, 2, \dots, p$, son constantes reales y el proceso $\{\varepsilon_t\}_t$ es ruido blanco con $\mathbb{E}(\varepsilon_t) = 0, \forall t$ y $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

La figura 2.6 ilustra una realización de un proceso AR(2).

Si se define el *operador de retraso* B como

$$BX_t = X_{t-1},$$

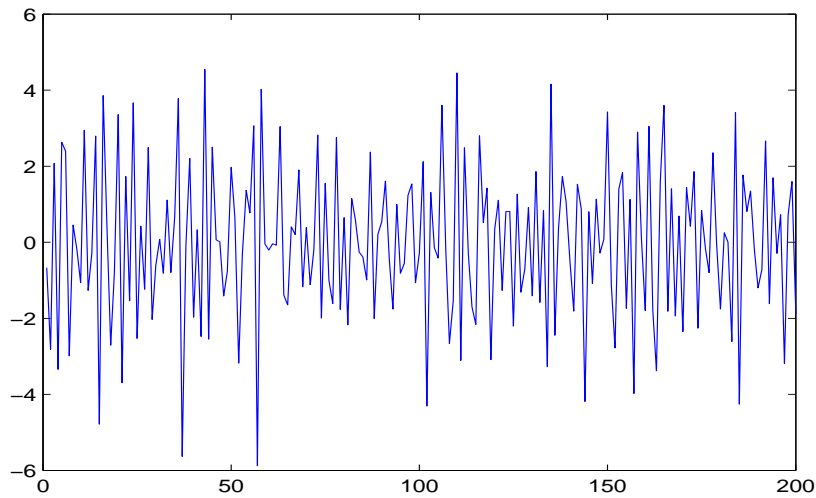


Figura 2.5: Una realización de un proceso MA(1)

entonces podemos escribir (2.4) como

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = \mu + \varepsilon_t. \quad (2.5)$$

La siguiente es una argumentación heurística para estudiar las condiciones bajo las cuales el modelo (2.4) es estacionario de 2º orden.

Pensemos en el modelo AR(1)

$$(1 - \phi B)X_t = \mu + \varepsilon_t;$$

nos hacemos la pregunta: ¿Cómo podemos encontrar $(1 - \phi B)^{-1}$ de forma que

$$(1 - \phi B)^{-1}(1 - \phi B)X_t = (1 - \phi B)^{-1}(\mu + \varepsilon_t)?$$

o equivalentemente, obtener al proceso $\{X_t\}_t$ en función del ruido $\{\varepsilon_t\}_t$

$$X_t = (1 - \phi B)^{-1}(\mu + \varepsilon_t).$$

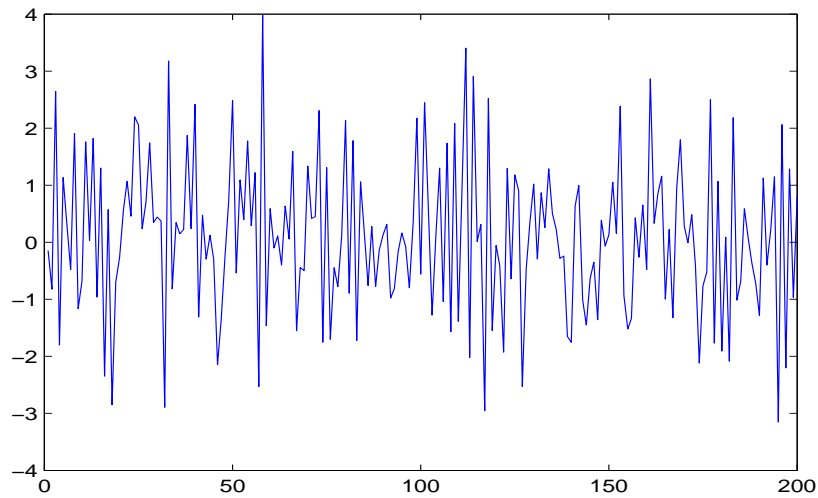


Figura 2.6: Ejemplo de un proceso AR(2)

Con el fin anterior pensemos que la “variable” B en el polinomio de retraso

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.6)$$

no es un operador sino el número complejo z , es decir, trabajaremos con el polinomio de variable compleja

$$\phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p. \quad (2.7)$$

Si pensamos en la operación composición $B^2 = B \circ B \equiv B \cdot B$ como “producto” para operadores podemos notar que las operaciones algebraicas que se pueden hacer con polinomios como (2.6) son iguales a las operaciones algebraicas que se pueden hacer con polinomios como (2.7), por ejemplo

$$(1 - z)(1 - z) = 1 - 2z + z^2$$

y por otra parte tenemos que

$$(1 - B)(1 - B) = (1 - 2B + B^2).$$

o bien

$$\begin{aligned}
(1 - B)(1 - B)X_t &= (1 - B)(X_t - X_{t-1}) \\
&= X_t - X_{t-1} - (X_{t-1} - X_{t-2}) \\
&= X_t - 2X_{t-1} + X_{t-2} \\
&= (1 - 2B + B^2)X_t.
\end{aligned}$$

Supuesto: Asumiendo que esta “igualdad de operaciones algebraicas” se cumple, entonces podemos proceder a estudiar la pregunta antes hecha, es decir ¿cómo encontrar $(1 - \phi B)^{-1}$ con polinomios del tipo (2.7)?.

Así entonces, para \mathbb{C} = conjunto de los números complejos sean

$$g(x) = 1 - \phi x \quad \text{y} \quad f(x) = \frac{1}{1 - \phi x},$$

donde $x \in \mathbb{R} \subset \mathbb{C}$ (aquí podemos pensar en los polinomios de retraso $g(B) = 1 - \phi B$ y $f(B) = (1 - \phi B)^{-1}$).

El polinomio $f(x) = (1 - \phi x)^{-1}$ tiene expansión en serie de Taylor (alrededor de $x = 0$) dada por:

$$(1 - \phi x)^{-1} = f(x) = 1 + \phi x + (\phi x)^2 + (\phi x)^3 + \dots = \sum_{k=0}^{\infty} (\phi x)^k,$$

siempre que $|\phi| < 1$.

De nuestras consideraciones anteriores, bajo este supuesto de igualdad en las operaciones algebraicas, tendríamos que $(1 - \phi B)^{-1} = \sum_{k=0}^{\infty} (\phi B)^k$.

Así, el proceso autorregresivo de orden $p = 1$ se puede escribir, para $|\phi| < 1$, como

$$\begin{aligned}
X_t &= (1 - \phi B)^{-1}(\mu + \varepsilon_t) \\
&= \sum_{k=0}^{\infty} (\phi B)^k(\mu + \varepsilon_t) \\
&= \frac{\mu}{1 - \phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots \\
&= \frac{\mu}{1 - \phi} + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k},
\end{aligned} \tag{2.8}$$

donde $\psi_k = \phi^k$. En la tercera igualdad asumimos que $B\mu = \mu$.

La igualdad (2.8) es una igualdad entre variables aleatorias y existen varios sentidos en los que ésta se puede entender.

Sea $L^2(\Omega, F, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{R} \mid X \text{ es variable aleatoria y } \mathbb{E}(X^2) < \infty\}$. Una de estas propiedades es que se puede definir una norma en $L^2(\Omega, F, \mathbb{P})$. Si $X \in L^2(\Omega, F, \mathbb{P})$

$$\|X\|_2 = \sqrt{\mathbb{E}[|X|^2]}$$

El operador B se puede pensar como un operador con dominio dado por un subconjunto de $L^2(\Omega, F, \mathbb{P})$ y con rango un subconjunto de $L^2(\Omega, F, \mathbb{P})$. La igualdad en (2.8) se debe entender como “igualdad en media cuadrática” es decir

$$\left\| X_t - \frac{\mu}{1-\phi} - \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \right\|_2 = \sqrt{\mathbb{E} \left[\left| X_t - \frac{\mu}{1-\phi} - \sum_{\psi=k}^{\infty} \psi_k \epsilon_{t-k} \right|^2 \right]} = 0.$$

Ahora, usando los teoremas de convergencia² y (2.8) concluimos que $\mathbb{E}(X_t) = \frac{\mu}{1-\phi} \equiv \mu_0$.

Podemos ahora considerar $Y_t = X_t - \mu_0$ y entonces

$$\begin{aligned} \text{Cov}[X_t, X_{t+\tau}] &= \mathbb{E}[Y_t Y_{t+\tau}] \\ &= \mathbb{E} \left[\left(\sum_{k=0}^{\infty} \psi_k \epsilon_{t-k} \right) \left(\sum_{l=0}^{\infty} \psi_l \epsilon_{t+\tau-l} \right) \right] \\ &= \mathbb{E} \left[\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \epsilon_{t-k} \epsilon_{t+\tau-l} \right] \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \mathbb{E}[\epsilon_{t-k} \epsilon_{t+\tau-l}]. \end{aligned}$$

La última igualdad se justifica usando los teoremas de convergencia dominada y/o monótona .

²Estos teoremas se detallarán en el Apéndice.

Por lo tanto

$$\text{Cov}[X_t, X_{t+\tau}] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi_k \psi_l \mathbb{E}[\varepsilon_{t-k} \varepsilon_{t+\tau-l}],$$

pero por ser $\{\varepsilon_t\}_t$ un proceso de ruido blanco,

$$\mathbb{E}[\varepsilon_{t-k} \varepsilon_{t+\tau-l}] = \begin{cases} 0, & \text{si } l \neq k + \tau, \\ \sigma_\varepsilon^2, & \text{si } l = k + \tau. \end{cases}$$

Luego

$$\begin{aligned} \text{Cov}[X_t, X_{t+\tau}] &= \sum_{k=0}^{\infty} \psi_k \psi_{k+\tau} \sigma_\varepsilon^2 \\ &= \sum_{k=0}^{\infty} \phi^{2k+\tau} \sigma_\varepsilon^2 \\ &= \phi^\tau \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \phi^{2k} \\ &= \sigma_\varepsilon^2 \left(\frac{\phi^\tau}{1-\phi^2} \right). \end{aligned} \tag{2.9}$$

Como $\mathbb{E}(X_t)$ y $\text{Cov}[X_t, X_{t+\tau}]$ no dependen de t , entonces $\{X_t\}_t$ es estacionario de 2º orden bajo el supuesto de que $|\phi| < 1$.

2.1.4. Proceso ARMA(p, q)

Diremos que el proceso a tiempo discreto X_t es tipo *Autorregresivo y Promedios Móviles* de orden p y q , ARMA(p, q), si es un proceso estacionario y para cada $t \in \{0, \pm 1, \pm 2, \dots\}$ se tiene

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_t, \tag{2.10}$$

donde $\{\varepsilon_t\}$ es un proceso de ruido blanco con $\mathbb{E}[\varepsilon_t] = 0$.

En la figura 2.7 se ilustra un ejemplo de un proceso ARMA(1,1).

Diremos que $\{X_t\}$ es un proceso ARMA(p, q) con media μ si $\{X_t - \mu\}$ es un proceso ARMA(p, q).

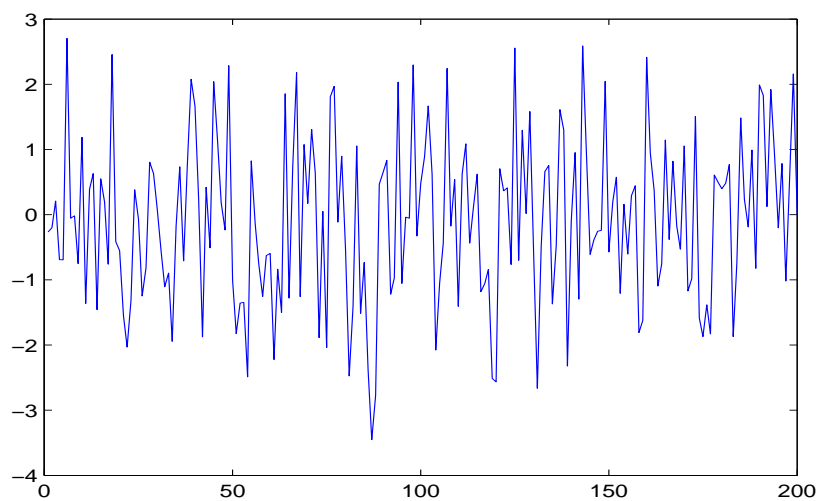


Figura 2.7: Una realización de un proceso ARMA(1,1).

La ecuación (2.10) se puede escribir usando la notación de operadores de retraso como

$$\phi_p(B)X_t = \theta_q(B)\varepsilon_t \quad ; \quad t \in \{0, \pm 1, \pm 2, \dots\},$$

con

$$\phi_p(z) = 1 - \phi_1 z - \dots - \phi_p z^p, \quad z \in \mathbb{C},$$

$$\theta_q(z) = 1 + \theta_1 z + \dots + \theta_q z^q, \quad z \in \mathbb{C},$$

y

$$B^j X_t = X_{t-j}.$$

El polinomio $\phi_p(z)$ se llama polinomio autoregresivo de orden p y el polinomio $\theta_q(z)$ se llama polinomio de promedios móviles de orden q .

Notemos que

(a) Si $\phi_p(z) \equiv 1$ entonces $X_t = \theta(B)\varepsilon_t$ es un modelo MA(q).

(b) Si $\theta_q(z) \equiv 1$ entonces $\phi(B)X_t = \varepsilon_t$ es un modelo AR(p).

Definición 7 (*Causalidad*)

Un proceso ARMA(p, q) definido por la ecuación (2.10) se llama causal (función causal de $\{\varepsilon_t\}$), si existe una sucesión de constantes $\{\psi_j\}_j$ tales que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.11)$$

La ecuación anterior se conoce en la literatura como *Modelo lineal general* o *Promedios móviles de orden $+\infty$* .

Notemos que, en (2.11), $\{X_t\}$ es el resultado de aplicarle a $\{\varepsilon_t\}$ un filtro **causal lineal**. En teoría de filtrado, una sucesión $\{g_u\}_u$ se llama filtro causal si en la expresión

$$X_t = \sum_{u=-\infty}^{\infty} g_u Y_{t-u},$$

se tiene que $g_u \equiv 0, \forall u < 0$. Es decir, para calcular X_t no se involucra información de la serie $\{Y_l\}_l$ para $l = t + 1, t + 2, \dots$ (es decir, información futura de $\{Y_t\}_t$).

El siguiente Teorema, cuya demostración puede consultarse en Brockwell y Davis (1987), establece condiciones necesarias y suficientes para que un proceso ARMA(p, q) sea causal.

Teorema 1

Sea $\{X_t\}_t$ un proceso ARMA(p, q) para el cual los polinomios $\phi_p(\cdot)$ y $\theta_q(\cdot)$ no tienen ceros en común. Entonces $\{X_t\}_t$ es un proceso causal si y sólo si $\phi_p(z) \neq 0, \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\psi_j\}_j$ en la ecuación

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \{0, \pm 1, \pm 2, \dots\},$$

quedan determinados por la relación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta_q(z)}{\phi_p(z)}, \quad |z| \leq 1. \quad (2.12)$$

A continuación, especificamos cómo calcular los pesos $\{\psi_j\}_j$.

Cálculo de los pesos $\{\psi_j\}_j$

Los valores de $\{\psi_j\}_j$ se pueden deducir de la relación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta_q(z)}{\phi_p(z)}; \quad |z| \leq 1, \quad (2.13)$$

donde

$$\theta_q(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \quad y$$

$$\phi_p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p.$$

Para lo subsecuente, convenimos lo siguiente:

$$\begin{aligned} \theta_0 &\equiv 1 \equiv \phi_0, \\ \phi_j &\equiv 0, \quad j > p, \\ \theta_j &\equiv 0, \quad j > q. \end{aligned} \quad (2.14)$$

La ecuación (2.13) se puede escribir como

$$\sum_{j=0}^{\infty} c_j z^j = \psi(z) \cdot \phi_p(z) = \theta_q(z) = \sum_{j=0}^{\infty} \theta_j z^j.$$

Entonces, debe suceder que para cada j

$$c_j = \sum_{k=0}^j (-\phi_k) \psi_{j-k} = \psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j, \quad (2.15)$$

donde, si usamos (2.14), obtenemos

$$c_0 = \psi_0 = \theta_0 = 1. \quad (2.16)$$

Ahora bien, en la ecuación (2.15), si se tiene $j \geq \max(p, q+1)$ entonces $j \geq p$ y $j \geq q+1$, luego $\theta_j = 0$ por (2.14) y (2.15) se escribe

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0.$$

De esta observación, la ecuación (2.15) tiene diferentes expresiones para dos dominios del valor de j

1. $\psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j ; 1 \leq j < \max(p, q+1),$
2. $\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0 ; j \geq \max(p, q+1).$

Podemos usar (1) y (2) para encontrar $\psi_j ; j = 1, 2, \dots$ en forma recursiva

$$\begin{aligned} \psi_0 &= \theta_0 = 1 \\ \psi_1 &= \theta_1 + \phi_1 \psi_0 = \theta_1 + \phi_1 \\ \psi_2 &= \theta_2 + \phi_1 \psi_1 + \phi_2 \psi_0 \\ &= \theta_2 + \phi_1 [\theta_1 + \phi_1] + \phi_2 \\ &= \phi_1^2 + \phi_2 + \phi_1 \theta_1 + \theta_2 \\ &\vdots \end{aligned} \tag{2.17}$$

Por ejemplo, sea $\{X_t\}_t$ un AR(1) con $|\phi_1| < 1$, entonces desarrollando

$$X_t = \phi_1 X_{t-1} + \varepsilon_t,$$

$$(1 - \phi_1 B)X_t = \varepsilon_t,$$

esto es, tenemos que $\phi(z) = 1 - \phi_1 z$ y $\theta(z) = 1 = \theta_0$ y haciendo las cuentas como en (2.17), $\psi_0 = \theta_0 = 1$, $\psi_1 = 0 + \phi_1 \cdot 1 = \phi_1$, $\psi_2 = 0 + \phi_1 \cdot \phi_1 + 0 \cdot \psi_0 = \phi_1^2$, etc.

Luego

$$X_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots$$

Un concepto muy relacionado con el de causalidad es el de invertibilidad.

Definición 8 (*Invertibilidad*)

Un proceso ARMA(p, q) definido por las ecuaciones

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad t \in \{0, \pm 1, \pm 2, \dots\},$$

se llama invertible, si existe una sucesión de constantes $\{\pi_j\}_j$ para la cual:

1. $\sum_{j=0}^{\infty} |\pi_j| < \infty$ y
2. $\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \{0, \pm 1, \pm 2, \dots\}.$

El siguiente teorema, cuya demostración puede verse en el libro de Brockwell y Davis (1987), da condiciones necesarias y suficientes para invertibilidad.

Teorema 2

Sea $\{X_t\}_t$ un proceso ARMA(p, q) para el cual los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no tienen ceros en común. Entonces $\{X_t\}_t$ es invertible si y sólo si $\theta(z) \neq 0, \forall z \in \bar{B}(0, 1)$. Los coeficientes $\{\pi_j\}_j$ en 1 y 2, quedan determinados por la relación

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1. \quad (2.18)$$

El cálculo de los pesos $\{\pi_j\}_j$ se puede llevar a cabo en forma análoga a como se hace para encontrar $\{\psi_j\}_j$ en la nota que antecede a la Definición 8.

La función de autocovarianza de un proceso ARMA(p, q) causal tiene el siguiente comportamiento asintótico.

Proposición 1

Sea $\{X_t\}_t$ un proceso ARMA(p, q) causal. Entonces si

$$\{S_h^X : h \in \{0, \pm 1, \pm 2, \dots\}\}$$

denota la sucesión de autocovarianza de $\{X_t\}_t$ se tiene que

$$\sum_{h=-\infty}^{\infty} |S_h^X| < \infty.$$

En particular $S_h^X \rightarrow 0$ cuando h tiende a ∞ para procesos $\{X_t\}_t$ ARMA(p, q) causales.

Estimación de la media del proceso estacionario de 2º orden

Sea $\{X_t\}_t$ un proceso estacionario de 2º orden. Un estimador natural para $\mu = \mathbb{E}[X_t]$ es

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

La media muestral es un estimador insesgado para μ , pues $\mathbb{E}[\bar{X}_n] = \mu$. En el siguiente resultado, se estudia el error cuadrático medio de \bar{X}_n , m.s.e.(n) $\equiv \mathbb{E}[(\bar{X}_n - \mu)^2]$.

Nuevamente las demostraciones de los siguientes resultados pueden encontrarse en el libro de Brockwell y Davis (1987).

Teorema 3

Si $\{X_t\}_t$ es un proceso estacionario de 2º orden con media μ y sucesión de autocovarianza $\{S_h^X\}_h$, entonces si $n \rightarrow \infty$

1. $\text{Var}(\bar{X}_n) = \mathbb{E}[(\bar{X}_n - \mu)^2] \rightarrow 0$, siempre que $S_h^X \rightarrow 0$, cuando $h \rightarrow \infty$.
2. $n\mathbb{E}[(\bar{X}_n - \mu)^2] \rightarrow \sum_{h=-\infty}^{\infty} S_h^X$, si $\sum_{h=-\infty}^{\infty} |S_h^X| < \infty$.

Así entonces el primer resultado del Teorema 3 establece que \bar{X} es un estimador consistente de $\mathbb{E}(X)$ para procesos estacionarios de 2º orden.

Cálculo de la función de autocovarianza

Primero determinaremos la función de autocovarianza, *autocovariance function* (ACVF) en inglés, $\{S_h^X\}_h$ de un proceso ARMA(p, q) causal definido por

$$\phi_p(B)X_t = \theta_q(B)\varepsilon_t, \quad \{\varepsilon_t\} \sim N(0, \sigma^2), \quad (2.19)$$

donde $\phi_p(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ y $\theta_q(z) = 1 + \theta_1 z + \dots + \theta_q z^q$. El supuesto de causalidad implica que

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

donde $\sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$, $|z| \leq 1$ y $\sum_{j=0}^{\infty} |\psi_j| < \infty$. En la literatura de series de tiempo existen varios métodos para realizar el cálculo que queremos efectuar; aquí solo presentaremos un método y se pueden encontrar más en Brockwell y Davis (1987).

Si multiplicamos cada lado de las ecuaciones

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

por X_{t-h} , $h = 0, 1, 2, \dots$, y tomamos esperanzas de cada lado, encontramos que

$$S_h^X - \phi_1 S_{h-1}^X - \dots - \phi_p S_{h-p}^X = \sigma^2 \sum_{j=0}^{\infty} \theta_{h+j} \psi_j, \quad 0 \leq h \leq m \quad (2.20)$$

y

$$S_h^X - \phi_1 S_{h-1}^X - \dots - \phi_p S_{h-p}^X = 0, \quad h \geq m, \quad (2.21)$$

donde $m = \max(p, q + 1)$, $\psi_j := 0$ para $j < 0$, $\theta_0 := 1$, y $\theta_j := 0$ para $j \notin \{0, \dots, q\}$.

Ejemplo 8 *Proceso ARMA(1,1)*

Para el proceso causal ARMA(1, 1) las ecuaciones anteriores son

$$S_0^X - \phi S_1^X = \sigma^2(1 + \theta(\theta + \phi)) \quad (2.22)$$

y

$$S_1^X - \phi S_0^X = \sigma^2\theta. \quad (2.23)$$

La ecuación (2.21) toma la forma

$$S_k^X - \phi S_{k-1}^X = 0, k \geq 2. \quad (2.24)$$

2.1.5. La función de autocovarianza y autocorrelación

Recalcamos que la función de autocorrelación (**ACF**) de un proceso AR-MA $\{X_t\}$ es la función ρ_h^X que se deduce inmediatamente de la función de autocovarianza S_h^X

$$\rho_h^X = \frac{S_h^X}{S_0^X}.$$

La función de autocovarianza (ACVF) y de autocorrelación muestrales, pueden ser obtenidas mediante las siguientes expresiones

$$\hat{S}_h^X = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}) \quad (2.25)$$

y

$$\hat{\rho}_h^X = \frac{\hat{S}_h^X}{\hat{S}_0^X}. \quad (2.26)$$

Los estimadores (2.25) y (2.26) no son insesgados pero cuando el tamaño de muestra $n \rightarrow \infty$ tienen distribución normal asintótica con media S_h^X y ρ_h^X respectivamente.

Así como expresamos la varianza de un modelo multivariado en forma matricial mediante la matriz de covarianzas, para las series de tiempo podemos encontrar una matriz de autocovarianzas que tiene la siguiente forma

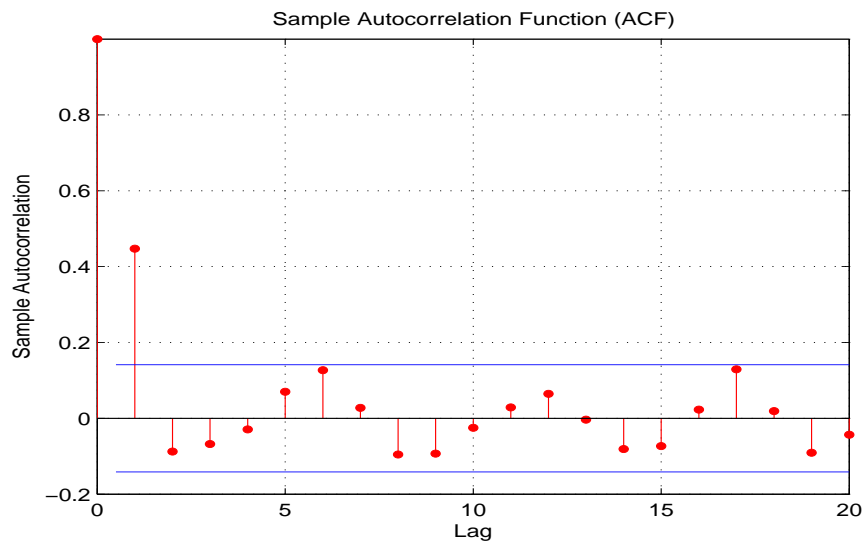


Figura 2.8: ACF muestral de un proceso MA(1)

$$\hat{\Sigma}_k = \begin{bmatrix} \hat{S}_0 & \hat{S}_1 & \dots & \hat{S}_{k-1} \\ \hat{S}_1 & \hat{S}_0 & \dots & \hat{S}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{S}_{k-1} & \hat{S}_{k-2} & \dots & \hat{S}_0 \end{bmatrix}.$$

Ahora vamos a discutir algunos métodos para calcular la función de autovarianza $\{S_h^X\}_h$ de un proceso ARMA causal $\{X_t\}_t$.

La ACF muestral de un proceso MA(q)

Dada unas observaciones $\{x_1, \dots, x_n\}$ de una serie de tiempo, una forma de diagnosticar un posible modelo para los datos es emparar la ACF muestral de los datos con la ACF del modelo. En particular, si la ACF muestral $\hat{\rho}_h^X$ es significativamente diferente de cero para $0 \leq h \leq q$ e insignificante para $h > q$, entonces esto sugiere que un modelo MA(q) puede proveer una buena representación de los datos (véase la forma de la ACVF teórica de un modelo MA(q) desarrollada en la sección 2.1.2).

Para aplicar este criterio necesitamos tomar en cuenta la variación aleatoria esperada en la función de autocorrelación muestral antes de que podamos clasificar los valores de la ACF como insignificantes. Una forma de resolver este problema es usar la fórmula de Bartlett³ la cual dice que para una muestra de tamaño n (grande) de un proceso MA(q), los valores de la ACF en los retrasos más grandes que q se distribuyen aproximadamente normal con media 0 y varianza $w_{hh}/n = (1 + 2\rho_1 + \dots + 2\rho_q)/n$. Esto significa que si una muestra proviene de un proceso MA(q) y si $h > q$, entonces $\hat{\rho}_h^X$ debe caer entre las bandas $\pm 1.96\sqrt{w_{hh}/n}$ con probabilidad 0.95 aproximadamente. En la práctica es frecuente el uso de los valores $\pm 1.96/\sqrt{n}$ como bandas entre las cuales la autocovarianza muestral es considerada insignificante.

En la figura 2.8 puede apreciarse una gráfica de la ACF de un proceso MA(1); podemos apreciar que sólo el primer retraso queda fuera de las bandas de confianza y todos los demás son insignificantes lo cual ilustra los conceptos anteriores.

2.1.6. La función de autocorrelación parcial

La función de autocorrelación parcial (PACF) de un proceso ARMA $\{X_t\}$ es la función $\alpha(\cdot)$ definida por las ecuaciones

$$\alpha(0) = 1$$

y

$$\alpha(h) = \phi_{hh}, h \geq 1,$$

donde ϕ_{hh} es el último componente del vector

$$\phi_h = \Sigma_h^{-1} \mathbf{S}_h, \quad (2.27)$$

$\Sigma_h = [S_{i-j}]_{i,j=1}^h$, y $\mathbf{S}_h = [S_1, S_2, \dots, S_h]'$. Para cada conjunto de observaciones $\{x_1, \dots, x_n\}$ la PACF muestral $\hat{\alpha}(h)$ está dada por

³Véase el libro de Brockwell y Davis (2002), capítulo 2, sección 2.4.2, pag. 61

$$\hat{\alpha}(0) = 1$$

y

$$\hat{\alpha}(h) = \hat{\phi}_{hh}, h \geq 1$$

donde $\hat{\phi}_{hh}$ es el último componente de

$$\hat{\phi}_h = \hat{\Sigma}_h^{-1} \hat{\mathbf{S}}_h, \quad (2.28)$$

donde $\hat{\Sigma}_h^{-1}$ y $\hat{\mathbf{S}}_h$ son los estimadores de Σ_h^{-1} y \mathbf{S}_h respectivamente.

La PACF muestral de un proceso AR(p)

Para los procesos AR(p), la PACF tiene un papel similar al papel que desempeña la ACF para los procesos MA(q). Puede probarse que si $\{X_t\}_t$ es un proceso AR(p), entonces $\alpha(h) = 0$ si $h > p$. Un argumento (más no una demostración) para justificar lo anterior es que para un proceso AR(p) causal definido por

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t, \quad \{\varepsilon_t\} \sim N(0, \sigma^2);$$

puede probarse que para $h \geq p$ el mejor predictor lineal de X_{h+1} en términos de $1, X_1, X_h$ es

$$\hat{X}_{h+1} = \phi_1 X_h + \phi_2 X_{h-1} + \cdots + \phi_p X_{h+1-p}.$$

Dado que el coeficiente ϕ_{hh} de X_h es ϕ_p si $h = p$ y 0 si $h > p$, concluimos que la PACF $\alpha(\cdot)$ del proceso $\{X_t\}$ tiene las propiedades

$$\alpha(p) = \phi_p$$

y

$$\alpha(h) = 0 \text{ para } h > p.$$

Para $h < p$ los valores de $\alpha(h)$ pueden ser calculados fácilmente de la ecuación (2.27).

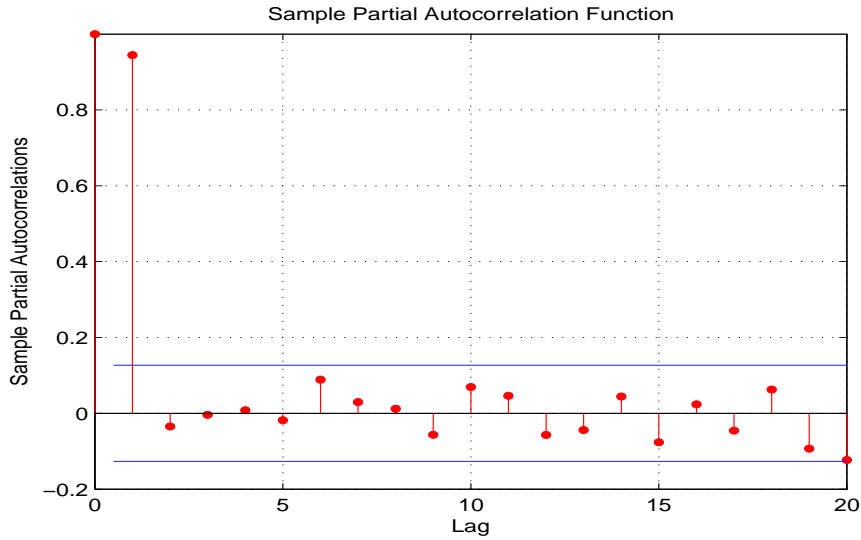


Figura 2.9: PACF muestral de un proceso AR(1)

De este comportamiento surge la siguiente idea de diagnóstico de un posible modelo $AR(p)$ para unos datos x_1, \dots, x_n . Dados estos datos calculamos $\{\hat{\alpha}(h)\}_h$ como en (2.28) y, al igual que en el caso de la ACF muestral $\hat{\rho}_h$, podemos determinar, mediante un gráfico, el valor p para el cual $\hat{\alpha}(h) \approx 0$ si $h > p$. Claramente esta no significancia se basa en un resultado estadístico. Los valores de la PACF muestrales en los retrasos más grandes que p son aproximadamente v.a. i.i.d. $N(0, 1/n)$. Esto significa que el 95% de los valores de la PACF muestral más allá del retraso p deben caer dentro de las bandas $\pm 1.96/\sqrt{n}$. Si observamos la PACF muestral satisface que $|\hat{\alpha}(h)| > 1.96/\sqrt{n}$ para $0 \leq h \leq p$ y $|\hat{\alpha}(h)| < 1.96/\sqrt{n}$ para $h > p$.

De esta forma, si se ha de seleccionar un modelo en la clase AR el orden propuesto usando $\{\hat{\alpha}(h)\}_h$ sería p . En la figura 2.9 se puede observar una gráfica de la PACF de un proceso AR(1); podemos ver cómo el primer retraso queda fuera de las bandas de confianza mientras que los retrasos sucesivos son despreciables.

2.2. Selección de modelos

Tanto la ACF como la PACF muestrales pueden ser utilizadas como herramientas para hacer un diagnóstico acerca de posibles modelos que puedan ajustarse a nuestro conjunto de datos. Como hemos visto, si nuestros datos siguen un proceso AR(p) la PACF puede servirnos para verificar este proceso y si se trata de un proceso MA(q) entonces la ACF nos servirá como evidencia.

Para modelos ARMA con $p > 0$ y $q > 0$, la ACF y PACF muestrales son difíciles de reconocer y son mucho menos efectivas para utilizarse como criterios para selección de modelos que en los casos en los que $p = 0$ y $q = 0$. Aun así puede utilizarse un criterio para selección de modelos a través de la minimización del *estadístico de AICC (Akaike Information Criterion Corrected)*

$$\text{AICC} = -2 \log L(\phi_p, \theta_q, S(\phi_p, \theta_q)/n) + 2(p + q + 1)n/(n - p - q - 2),$$

el cual será discutido más adelante. Para p y q fijos es claro de la definición que el valor del AICC es minimizado por lo valores de los parámetros que maximizan la verosimilitud. Por tanto, la decisión final para considerar el orden p y q que minimiza el AICC debe ser basado en la estimación por máxima verosimilitud que se describirá a continuación.

2.2.1. Estimación de parámetros por máxima verosimilitud

Considerese $t^* = \max(p, q)$ y un tamaño de muestra N . Nuevamente en Brockwell y Davis (1987) puede consultarse que la función de verosimilitud para un conjunto de datos provenientes de un procesos ARMA(p, q) tiene la forma siguiente

$$L(\Phi_p, \Theta_q, \sigma_\varepsilon^2) = f(X_{t^*+1}, \dots, X_N | \Phi_p, \Theta_q, \sigma_\varepsilon^2) \quad (2.29)$$

$$= (2\pi)^{-\frac{(N-t^*)}{2}} \sigma^{-(N-t^*)} e^{-\sum_{t=t^*+1}^N S(\Phi_p, \Theta_q)/2\sigma_\varepsilon^2} \cdot 1 \quad (2.30)$$

donde $S(\Phi_{\mathbf{p}}, \Theta_{\mathbf{q}}) = \sum_{t=t^*+1}^N (X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})^2$,

$$\Phi_{\mathbf{p}} = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}, \quad \Theta_{\mathbf{q}} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{pmatrix},$$

Dada σ_ε^2 fija, para minimizar L como función de Φ y Θ , hay que minimizar $S(\Phi_{\mathbf{p}}, \Theta_{\mathbf{q}})$. Lo anterior permitirá maximizar la expresión $e^{-S(\Phi_{\mathbf{p}}, \Theta_{\mathbf{q}})}$ pues, al ser $S(\Phi_{\mathbf{p}}, \Theta_{\mathbf{q}})$ una suma de cuadrados, al igual que en el análisis de regresión, al minimizar la suma de cuadrados se maximiza la verosimilitud.

Supóngase que se encuentran $\hat{\Phi}_{\mathbf{p}}$ y $\hat{\Theta}_{\mathbf{q}}$ que minimizan $S(\Phi_{\mathbf{p}}, \Theta_{\mathbf{q}})$; entonces se procede a maximizar (2.29) con respecto a σ_ε^2 . Equivalentemente pensemos en maximizar

$$\log L(\sigma_\varepsilon^2, \hat{\Phi}_{\mathbf{p}}, \hat{\Theta}_{\mathbf{q}}) = \left(\frac{-N - t^*}{2} \right) [\log(2\pi) + \log(\sigma_\varepsilon^2)] - S(\hat{\Phi}_{\mathbf{p}}, \hat{\Theta}_{\mathbf{q}})/2\sigma_\varepsilon^2.$$

con respecto a σ_ε^2 . El estimador para σ_ε^2 posee una solución analítica cerrada, dada por

$$\begin{aligned} \frac{\partial \log L}{\partial \sigma_\varepsilon^2} &= (-N - t^*)/2\sigma_\varepsilon^2 + S(\hat{\Phi}_{\mathbf{p}}, \hat{\Theta}_{\mathbf{q}})/2\sigma_\varepsilon^4 = 0 \\ \frac{\partial \log L}{\partial \sigma_\varepsilon^2} |_{\sigma_\varepsilon^2 = \hat{\sigma}_\varepsilon^2} &= \hat{\sigma}_\varepsilon^2 = 0 \iff \hat{\sigma}_\varepsilon^2 = S(\hat{\Phi}_{\mathbf{p}}, \hat{\Theta}_{\mathbf{q}})/(N - t^*). \end{aligned}$$

Como se puede apreciar, las derivadas parciales con respecto a los vectores $\hat{\Phi}_{\mathbf{p}}$ y $\hat{\Theta}_{\mathbf{q}}$ no tienen una solución analítica cerrada, por lo cual deben utilizarse métodos de optimización numérica para maximizar la log-verosimilitud. Cabe mencionar que para poder implementar la solución de manera numérica se requieren soluciones preeliminares para los vectores $\hat{\Phi}_{\mathbf{p}}$ y $\hat{\Theta}_{\mathbf{q}}$ y éstas pueden obtenerse mediante diversos métodos. Una manera de hacer ésto es estimar por separado el modelo como si fuera un proceso AR(p) con las *ecuaciones de Yule-Walker* y por otra parte estimar un proceso MA(q) con el *Algoritmo de Innovación* (véase Brockwell y Davis, 2000).

Anteriormente mencionamos que un método para seleccionar modelos se tiene al minimizar el valor del estadístico AICC. Es el mejor criterio para seleccionar los órdenes de p y q . Este método se aplica como sigue (véase Brockwell y Davis 2000 pág. 161, Shumway y Stoffer 2006 pág. 54):

Definición 9 *Selección de modelos usando el estadístico de información de Akaike corregida (AICC)*

Escoger p , q , $\hat{\Phi}_p$ y $\hat{\Theta}_q$ tales que

$$\text{AICC} = -2 \log L(\hat{\Phi}_p, \hat{\Theta}_q, S(\hat{\Phi}_p, \hat{\Theta}_q)/n) + 2(p + q + 1)n/(n - p - q - 2)$$

sea mínimo.

Como habíamos visto anteriormente, para p y q fijos es claro que el AICC es minimizado cuando $\hat{\Phi}_p$ y $\hat{\Theta}_q$ son los vectores que minimizan $-2 \log L(\hat{\Phi}_p, \hat{\Theta}_q, S(\hat{\Phi}_p, \hat{\Theta}_q)/n)$, es decir, los estimadores máximo verosímiles. De igual manera la decisión final con respecto la selección de orden debe hacerse en base a los estimadores de máxima verosimilitud en vez de tomar como referencia los estimadores iniciales que sólo sirven como una guía para encontrar los de máxima verosimilitud.

2.2.2. Análisis de Residuales

Una vez que hemos estimado los parámetros, lo siguiente es verificar la bondad de ajuste del modelo con respecto a nuestras observaciones. Por lo cual la siguiente etapa en la modelación de datos es realizar el *análisis de residuales* correspondiente. La bondad de ajuste del modelo estadístico se juzga al comparar los valores observados con las predicciones obtenidas a través del modelo ajustado. Si el modelo es apropiado entonces los residuales deberán mantener un comportamiento consistente con el modelo. Por ello es necesario verificar que se cumplen los supuestos probabilísticos hechos inicialmente acerca de nuestro proceso de ruido blanco. Recordemos que en un modelo ARMA(p, q) la sucesión $\{\varepsilon_t\}_t \sim N(0, \sigma_\varepsilon^2)$; entonces definimos a los residuales como la diferencia entre el valor observado X_t y el valor de predicción $\hat{X}_t(\hat{\Phi}_p, \hat{\Theta}_q)$ obtenido usando el modelo, de la siguiente manera

$$\widehat{W}_t \equiv \frac{X_t - \hat{X}_t(\hat{\phi}_{\mathbf{p}}, \hat{\theta}_{\mathbf{q}})}{\sqrt{\hat{r}_{t-1}(\hat{\Phi}_{\mathbf{p}}, \hat{\Theta}_{\mathbf{q}})}}$$

donde $\hat{r}_{t-1} = \mathbb{E}[(X_t - \hat{X}_t)^2] = \mathbb{E}[\varepsilon_t^2] = \hat{\sigma}_\varepsilon^2$.

Entonces \widehat{W}_t debe poseer características probabilísticas similares a ε_t . Luego es necesario realizar una serie de pruebas estadísticas sobre la sucesión $\{\widehat{W}_t\}_t$ para verificar cada uno de los supuestos acerca del modelo, entre las que destacan las siguientes:

(a) Estacionaridad.

Una gráfica de los residuales $\{\widehat{W}_t\}_t$ vista como una serie de tiempo es útil para verificar el supuesto de estacionaridad el proceso. Si la gráfica presenta alguna tendencia o algún componente de periodicidad entonces no podemos afirmar que estamos trabajando con datos provenientes de un proceso estacionario.

(b) No-correlación.

Uno de los supuestos más importantes que deben de cumplir los residuales es la no-correlación. Una manera directa de verificar esto es con la autocorrelación muestral a través de la cantidad

$$\hat{\rho}_h^W = \frac{\sum_{t=1}^{n-h} (\widehat{W}_t - \bar{W})(\widehat{W}_{t+h} - \bar{W})}{\sum_{t=1}^n (\widehat{W}_t - \bar{W})^2},$$

$$\bar{W} = \frac{1}{n} \sum_{t=1}^n \widehat{W}_t$$

Un método alternativo es realizar un gráfico de la ACF muestral de los residuales. Esto nos ayudará a verificar el supuesto de no-correlación en \widehat{W}_t por lo cual en ningún retraso se deben de sobrepasar las bandas de confianza de la gráfica. En ocasiones la evaluación de los términos de la función de autocorrelación puede resultar insuficiente para diagnósticos. Un análisis más riguroso puede consistir en aplicar alguna prueba estadística como es la prueba de *Box-Ljung* (véase Shumway y Stoffer

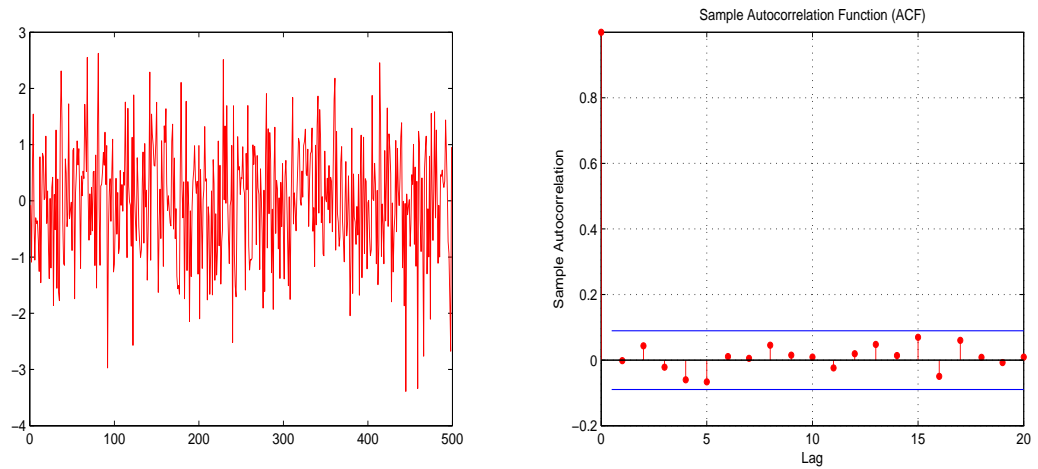


Figura 2.10: Gráfica y ACF de los residuales

2006 pág. 149) la cual nos indica si hay autocorrelación en un conjunto de datos.

(c) Prueba de bondad de ajuste.

Posteriormente, una de las pruebas vitales a realizar es la bondad de ajuste de los residuales con respecto a la distribución del proceso de ruido blanco. Como estamos asumiendo normalidad en la sucesión \widehat{W}_t los residuales deben de poseer esta misma distribución; un histograma puede ayudarnos a verificar esta hipótesis de una manera visual pero no es el método más eficiente. Un método gráfico que es de mayor utilidad es la gráfica de cuantil-cuantil ó *QQ-plot* en la cual aparecen los datos graficados contra los cuantiles teóricos que le corresponderían si éstos provinieran de una distribución normal, con lo cual este gráfico debe ser muy similar a una recta de 45 grados. Además también nos ayuda a visualizar algunas otras características de la distribución de nuestros datos como la presencia de asimetría. Aun así, éste no es un método tan formal comparado con una prueba de hipótesis por lo que es recomendable realizar una prueba de normalidad como la prueba de *Jarque-Bera* o la prueba de *Kolmogorov-Smirnov*. En el último capítulo describiremos estas pruebas e ilustraremos la manera en que se realiza

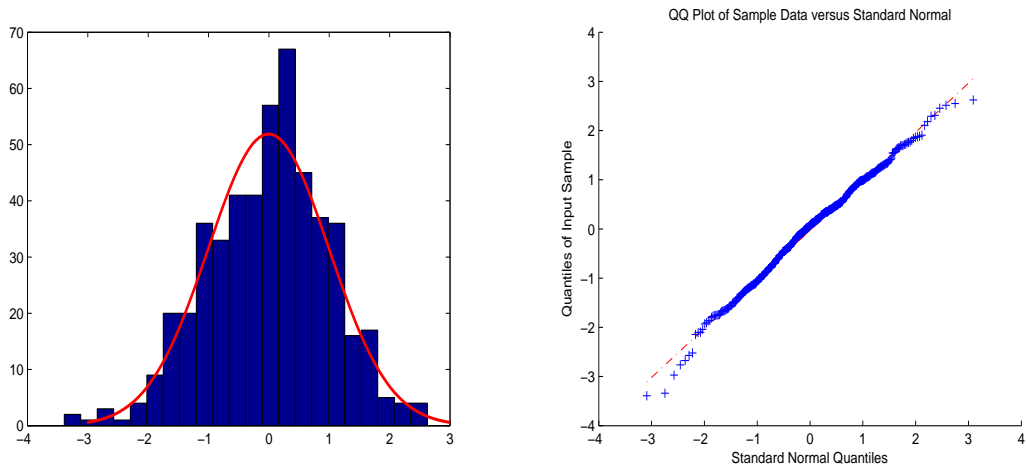


Figura 2.11: Histograma y QQ-plot de los residuales

el análisis de residuales siguiendo los pasos ya mencionados.

2.2.3. Predicción

La etapa final al construir un modelo de series de tiempo, una vez que se ha verificado la bondad del mismo, es poder hacer predicción sobre alguna o algunas observaciones futuras. En particular nos centraremos en la manera de hacer predicción en modelos $AR(p)$. Puede probarse que, para un conjunto de observaciones X_1, X_2, \dots, X_n que se pueden ajustar mediante un modelo $AR(p)$, el mejor predictor lineal tiene la forma

$$\hat{X}_{n+1} = \hat{\phi}_1 X_n + \hat{\phi}_2 X_{n-1} + \dots + \hat{\phi}_p X_{n-p+1}$$

donde \hat{X}_{n+1} es la observación futura cuyo valor se desea predecir. Comúnmente se acompaña la predicción media con un intervalo de confianza; ésto se describirá con mayor detalle nuevamente en el capítulo final. Cabe mencionar que si realizamos predicciones mediante un modelo $AR(p)$, sólo las primeras p predicciones podrán ser buenos estimadores para las observaciones futuras.

Si se desea realizar un número mayor de predicciones es conveniente esperar a contar con nuevas observaciones y reajustar el modelo.

2.3. Series de tiempo no lineales

En contraste con los modelos tradicionales de análisis de series de tiempo, los cuales se concentran en modelar la media condicional del proceso, en los modelos ARCH y GARCH se propone considerar la dependencia en las observaciones pasadas de la varianza condicional dentro del modelo. Las ecuaciones para la media no toman en cuenta los efectos de la heterocedasticidad como el apilamiento de la volatilidad (*volatility clustering*) en algunas series de tiempo observadas o las colas pesadas en la distribución de X_t , pues estas características aparecen en datos que se estudian en el campo de las finanzas.

2.3.1. Propiedades básicas del proceso ARCH(q)

Definición 10 (*Proceso ARCH*)

Un modelo autorregresivo condicionalmente heterocedástico (ARCH) con orden $q(\geq 1)$ es definido como

$$X_t = \sigma_t \varepsilon_t, \quad \text{donde} \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_q X_{t-q}^2, \quad (2.31)$$

para constantes $\alpha_0 \geq 0, \alpha_j \geq 0$ y $\{\varepsilon_t\}_t$ v.a.'s i.i.d. con distribución \mathcal{D}_ϑ , cuya media es 0 y cuya varianza es 1. Se asume que ε_t es independiente de $\{X_{t-k}, k \geq 1\}$ para toda t . Un proceso estocástico $\{X_t\}$ definido por las ecuaciones de arriba es llamado *proceso ARCH(q)*.

El modelo ARCH fue introducido por Engle (1982) para modelar y predecir la varianza para las tasas de inflación de Reino Unido. Desde entonces, ha sido ampliamente usado para modelar la volatilidad de series de tiempo económicas y financieras. La idea básica en la construcción de (2.31) es muy intuitiva: la distribución predictiva de X_t se obtiene de reescalar una distribución con varianza 1 (la de $\{\varepsilon_t\}_t$) con una constante σ_t que depende del pasado del proceso. Por lo tanto, los cuantiles condicionales de X_t dado su pasado, los cuales juegan un rol muy importante en la administración de

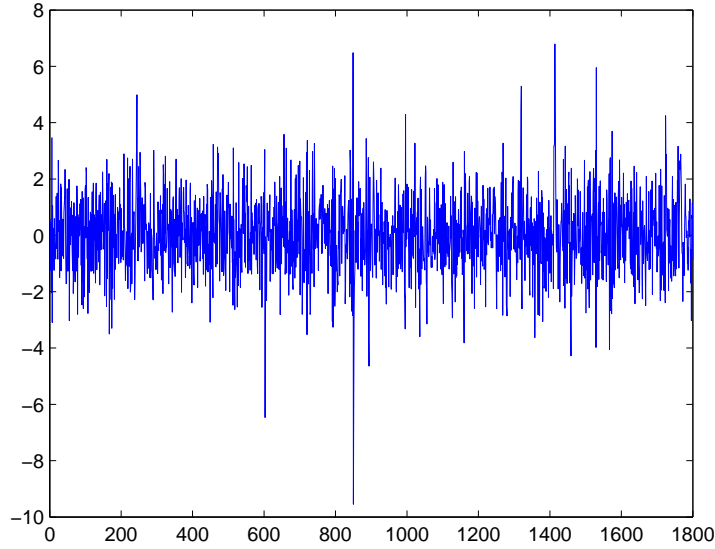


Figura 2.12: Realización de un proceso ARCH(1)

riesgos financieros (se detallará en el capítulo 4), pueden ser evaluados fácilmente. Por ejemplo si $\varepsilon_t \sim N(0, 1)$, la distribución predictiva es $N(0, \sigma_t^2)$. Esto marca un contraste con los modelos de series de tiempo lineales en los cuales el error cuadrático medio condicional es constante. Una realización de este proceso de muestra en la figura 2.12

Para entender el modelo ARCH, debemos poner atención en especial al modelo ARCH(1)

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2,$$

donde $\alpha_0 > 0$ y $\alpha_1 \geq 0$. Primero, la media de X_t se reduce a cero pues

$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{F}_{t-1}]] = \mathbb{E}[\sigma_t \mathbb{E}[\varepsilon_t]] = 0 \quad (2.32)$$

donde \mathcal{F}_{t-1} denota la historia del proceso $\{X_t\}_t$ hasta el tiempo $t - 1$.

Segundo, la varianza de X_t puede ser obtenida como

$$\begin{aligned}
\text{Var}(X_t) &= \mathbb{E}[X_t^2] = \mathbb{E}[\mathbb{E}[X_t^2|\mathcal{F}_{t-1}]] = \mathbb{E}[\mathbb{E}[X_t^2|X_{t-1}, \dots, X_0]] \\
&= \mathbb{E}[\sigma_t^2 \varepsilon_t^2] = \mathbb{E}[\sigma_t^2] \mathbb{E}[\varepsilon_t^2] = \mathbb{E}[\alpha_0 + \alpha_1 X_{t-1}^2] \\
&= \alpha_0 + \alpha_1 \mathbb{E}[X_{t-1}^2] = \alpha_0 + \alpha_1 \text{Var}(X_{t-1})
\end{aligned}$$

Si asumimos que X_t es un proceso estacionario con $\mathbb{E}[X_t] = 0$, entonces $\text{Var}(X_t) = \text{Var}(X_{t-1}) = \mathbb{E}(X_{t-1}^2)$. Por lo tanto $\text{Var}(X_t) = \alpha_0 + \alpha_1 \text{Var}(X_t)$ y entonces $\text{Var}(X_t) = \alpha_0 / (1 - \alpha_1)$. Como la varianza de X_t debe ser positiva, se necesita que $0 < \alpha_1 < 1$.

Tercero, en algunas aplicaciones, necesitamos la existencia de momentos de orden más alto por lo que α_1 también debe satisfacer la restricción $0 < \alpha_1 < 1$. Por lo pronto, para estudiar el comportamiento de las colas de la distribución de X_t , requerimos que el cuarto momento de $\{X_t\}_t$ sea finito. Bajo supuesto de normalidad para ε_t , tenemos que

$$\mathbb{E}(X_t^4|\mathcal{F}_{t-1}) = 3[\mathbb{E}(X_t^2|\mathcal{F}_{t-1})]^2 = 3(\alpha_0 + \alpha_1 X_{t-1}^2)^2.$$

Luego

$$\mathbb{E}(X_t^4) = \mathbb{E}[\mathbb{E}(X_t^4|\mathcal{F}_{t-1})] = 3\mathbb{E}(\alpha_0 + \alpha_1 X_{t-1}^2)^2 = 3\mathbb{E}[\alpha_0^2 + 2\alpha_0\alpha_1 X_{t-1}^2 + \alpha_1^2 X_{t-1}^4].$$

Nuevamente, asumiendo que X_t es estacionario de orden cuarto con $m_4 = \mathbb{E}(X_t^4)$, entonces tenemos que

$$\begin{aligned}
m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(X_t) + \alpha_1^2 m_4] \\
&= 3\alpha_0 \left(1 + 2\left(\frac{\alpha_1}{1 - \alpha_1}\right) \right) + 3\alpha_1^2 m_4.
\end{aligned}$$

Consecuentemente,

$$m_4 = \frac{3\alpha_0(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}.$$

Este resultado tiene dos implicaciones importantes:

- (a) Dado que el cuarto momento de X_t es positivo, tenemos que α_1 debe satisfacer la condición $1 - 3\alpha_1^2 > 0$; esto es, $0 \leq \alpha_1^2 \leq \frac{1}{3}$;
- (b) La curtosis de X_t es

$$\frac{\mathbb{E}(X_t^4)}{[\text{Var}(X_t)]^2} = 3 \left(\frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \right) \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \left(\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right) > 3.$$

Así, la distribución de X_t tendrá mayor curtosis que una densidad Normal y por ende las colas de la distribución de X_t serán más pesadas que las de una v.a. Normal. Estas propiedades se cumplen en general para modelos ARCH, pero las fórmulas se vuelven más complicadas para modelos ARCH de orden mayor. El siguiente teorema (véase Fan y Yao, 2002, pág. 144) establece las condiciones necesarias y suficientes para que un proceso ARCH(q) sea estacionario.

Teorema 4

- (i) La condición necesaria y suficiente para que el proceso $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ en (2.31) sea definido como un proceso estrictamente estacionario con $\mathbb{E}[X_t^2] < \infty$ es $\sum_{j=1}^q \alpha_j < 1$. Además, bajo este supuesto

$$\mathbb{E}[X_t] = 0 \quad y \quad \mathbb{E}[X_t^2] = \alpha_0 / \left\{ 1 - \sum_{j=1}^q \alpha_j \right\},$$

y $X_t \equiv 0$ para toda t si $\alpha_0 = 0$.

- (ii) Si $\mathbb{E}[\varepsilon_t^4] < \infty$ y

$$\text{máx} \{1, (\mathbb{E}[\varepsilon_t^4])^{\frac{1}{2}}\} \times \sum_{j=1}^q \alpha_j < 1.$$

Entonces (2.31) tiene una solución estrictamente estacionaria con cuarto momento finito, nombrado $\mathbb{E}[X_t^4] < \infty$.

Desventajas de los modelos ARCH

Las ventajas de los modelos ARCH incluyen las propiedades discutidas previamente. El modelo también tiene algunas desventajas (véase Tsay, 2002, pág. 86) :

1. El modelo asume que las observaciones tanto positivas como negativas tienen los mismos efectos sobre la volatilidad debido a que dependen de los cuadrados de las observaciones previas. En la práctica se sabe que los precios de los activos financieros responden diferentemente a impactos positivos o negativos en la volatilidad.
2. El modelo ARCH es más bien restrictivo. Por ejemplo, para un modelo ARCH(1) α_1^2 debe estar en el intervalo $[0, \frac{1}{3}]$ si es que la serie tiene cuarto momento finito. La restricción se vuelve más complicada para modelos ARCH de orden mayor.
3. El modelo ARCH no ofrece una nueva visión para entender la fuente de las variaciones de una serie de tiempo financiera. Sólo proporcionan una manera mecánica de describir el comportamiento de la varianza condicional. No da ninguna indicación sobre las causas que originan tal comportamiento en los datos.
4. El modelo ARCH tiende a predecir valores más altos para la volatilidad debido a que responde lentamente a grandes impactos aislados para la serie de retornos⁴

Diagnóstico y selección de orden del modelo

Si utilizamos la PACF de X_t^2 podemos determinar el orden q del modelo ARCH. Eso se justificará a continuación. Retomando la ecuación (2.31) la volatilidad tiene la siguiente forma

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_q X_{t-q}^2.$$

⁴Definimos al retorno de un activo financiero como la variación porcentual del precio actual con respecto al precio del día anterior. De esta manera, si X_t denota el precio de un activo al tiempo t , el retorno será $Y_t = (X_t - X_{t-1})/X_{t-1}$; en ocasiones es más común trabajar con los *log-retornos*, es decir, $Y_t = \log(X_t/X_{t-1})$. Esta última expresión será la que usaremos como definición de retorno.

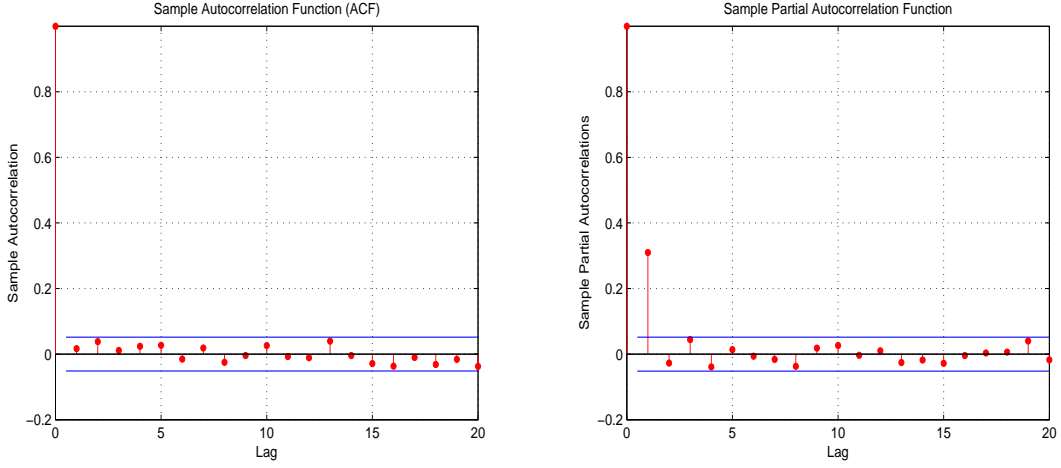


Figura 2.13: ACF de la series de datos y PACF de la serie al cuadrado

Para una muestra dada, X_t^2 es un estimador insesgado de σ_t^2 . Por tanto, esperamos que X_t^2 esté linealmente relacionado con $X_{t-1}^2, \dots, X_{t-q}^2$ de manera similar que un modelo autorregresivo de orden q . Alternativamente, definamos $\eta_t = X_t^2 - \sigma_t^2$. Se puede mostrar que $\{\eta_t\}$ es una serie no-correlacionada con media 0, es decir $\{\eta_t\}_t$ tiene estructura de ruido blanco. El modelo ARCH se convierte entonces en

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_q X_{t-q}^2 + \eta_t, \quad (2.33)$$

el cual tiene la forma de un modelo AR(q) para X_t^2 . Como se mencionó en las secciones previas, la PACF de X_t^2 es una herramienta útil para determinar el orden q . Como $\{\eta_t\}$ no es distribuída idénticamente, la estimación por mínimos cuadrados de (2.33) es *consistente* pero no *eficiente*. La PACF de X_t^2 puede no ser efectiva cuando el tamaño de muestra es pequeño. En la figura 2.13 se puede ver un gráfico de la ACF de una serie de tiempo donde vemos que no hay autocorrelación. En la parte derecha de la figura vemos la PACF de los cuadrados de la serie, donde vemos que el primer retraso queda fuera de las bandas de confianzas, lo cual nos sugiere que las observaciones pueden provenir de un modelo ARCH(1).

Para un modelo ARCH, los residuales estandarizados

$$\hat{\varepsilon}_t = X_t / \hat{\sigma}_t,$$

$$\hat{\sigma}_t = \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1}^2 + \cdots + \hat{\alpha}_q X_{t-q}^2}$$

deberían estar cercanos a ser v.a.'s i.i.d. con distribución normal \mathcal{D}_ϑ según se haya tomado en la definición (2.31). Por tanto, uno puede checar la bondad de ajuste del modelo ARCH al examinar la serie $\{\hat{\varepsilon}_t\}$. El sesgo, curtosis y el QQ-plot de $\{\hat{\varepsilon}_t\}$ pueden ser usados para verificar la validez de los supuestos distribucionales.

Las predicciones del modelo ARCH en la ecuación (2.31) pueden ser obtenidas recursivamente como en un modelo AR. Consideremos un modelo ARCH(q); la predicción para una observación futura σ_{n+1} tiene la siguiente forma (ver Tsay, 2002, págs. 89 y 90)

$$\sigma_n^2(1) = \alpha_0 + \alpha_1 X_n^2 + \cdots + \alpha_q X_{n+1-q}^2.$$

La predicción para una observación futura con 2 horizontes de tiempo sería

$$\sigma_n^2(2) = \alpha_0 + \alpha_1 \sigma_n^2(1) + \alpha_2 X_n^2 + \cdots + \alpha_q X_{n+2-q}^2.$$

y la predicción al horizonte de tiempo k para σ_{n+k} es

$$\sigma_n^2(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \sigma_n^2(k-i)$$

donde $\sigma_n^2(k-i) = X_{n+k-i}^2$ si $k-i \leq 0$.

2.3.2. Propiedades básicas del proceso GARCH(p, q)

Aunque el modelo ARCH es sencillo, regularmente se requieren más parámetros para describir adecuadamente el proceso de volatilidad de los rendimientos de algún activo financiero. Es por ello que el modelo ARCH ha sido extendido en varias direcciones, algunos hacia consideraciones económicas, otros a ideas estadísticas. La más importante de éstas es la extensión para incluir un

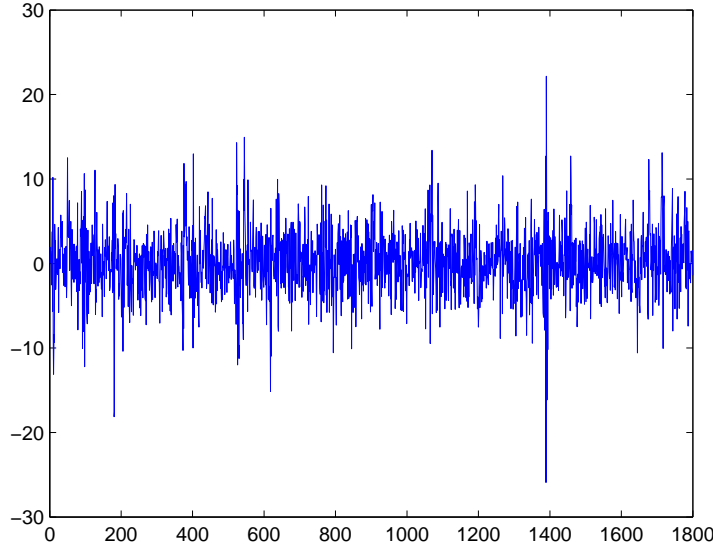


Figura 2.14: Realización de un proceso GARCH(1,1)

componente autorregresivo en la varianza condicional; esta extensión se conoce como *modelo ARCH generalizado (GARCH)* debido a Bollerslev (1986) y Taylor (1986). El proceso $\{X_t\}_t$ sigue un proceso GARCH(p, q) si

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (2.34)$$

donde nuevamente $\{\varepsilon_t\}$ es una sucesión de v.a.'s i.i.d. con media 0 y varianza unitaria, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, y $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$. Aquí se entiende que $\alpha_i = 0$ para $i > q$ y $\beta_j = 0$ para $j > p$. La restricción sobre $\alpha_i + \beta_i$ implica que la varianza de X_t es finita, mientras que la varianza condicional $\sigma_t^2 = \text{Var}(X_t | X_{t-1}, \dots, X_{t-q}, \sigma_{t-1}, \dots, \sigma_{t-p})$ evoluciona sobre el tiempo. Como se mencionó anteriormente, se asume que ε_t tiene una distribución \mathcal{D}_ϑ . La ecuación (2.34) se reduce a un modelo ARCH(q) puro si $p = 0$. Para entender las propiedades del modelo GARCH, es más conveniente utilizar la siguiente representación.

Sea $\nu_t = X_t^2 - \sigma_t^2$ tal que $\sigma_t^2 = X_t^2 - \nu_t$. Al agregar $\sigma_{t-i}^2 = X_{t-i}^2 - \nu_{t-i}$ ($i = 0, \dots, p$) dentro de (2.34), podemos reescribir el modelo GARCH como

$$X_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 + \nu_t - \sum_{j=1}^p \beta_j \nu_{t-j}. \quad (2.35)$$

Es fácil ver que $\mathbb{E}(\nu_t) = 0$ y $\text{Cov}(\nu_t, \nu_{t-j}) = 0$ para $j \geq 1$. De cualquier manera, $\{\nu_t\}$ en general no es una sucesión iid. La ecuación (2.35) es de la forma de un modelo ARMA para los cuadrados de la serie X_t^2 . Así un modelo GARCH puede ser visto como un modelo ARMA para los cuadrados de la serie X_t^2 . Usando la media de un modelo ARMA, tenemos que

$$\mathbb{E}(X_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)}$$

asumiendo que el denominador de la fracción anterior es positivo. Un ejemplo de este proceso puede apreciarse en la gráfica 2.14.

Las ventajas y desventajas del modelo GARCH pueden ser vistas fácilmente al concentrarnos en el modelo más sencillo de este género, el modelo GARCH(1,1), en el cual

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad 0 \leq \alpha_1, \beta_1 \leq 1, (\alpha_1 + \beta_1) < 1. \quad (2.36)$$

Primero, valores muy grandes de X_{t-1}^2 ó σ_{t-1}^2 producen un valor muy grande para σ_t^2 . Esto significa que un valor grande de X_{t-1}^2 tiende a ser seguido de otro valor grande de X_t^2 , generando, nuevamente, el bien conocido comportamiento de *volatility clustering* en series de tiempo financieras. Segundo, puede mostrarse que si $2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, entonces

$$\frac{\mathbb{E}(X_t^4)}{[\mathbb{E}(X_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

Consecuentemente, similar a los modelos ARCH, la cola de la distribución de un procesos GARCH(1, 1) es más pesada que la de una distribución normal. Tercero, el modelo provee una función paramétrica sencilla que puede usarse para describir la evolución de la volatilidad.

Las predicciones de un modelo GARCH pueden ser obtenidas usando métodos similares a los usados en un modelo ARMA. Considerese un modelo GARCH(1, 1) como en la ecuación(2.36) y asumamos que contamos con un conjunto de observciones de tamaño n . Para la predicción a un paso, tenemos

$$\sigma_{n+1}^2 = \alpha_0 + \alpha_1 X_n^2 + \beta_1 \sigma_n^2,$$

donde X_n y σ_n^2 son conocidos al tiempo n . Por lo tanto, la predicción a un paso es

$$\sigma_n^2(1) = \alpha_0 + \alpha_1 X_n^2 + \beta_1 \sigma_n^2.$$

Para predicciones múltiples, usamos $X_t^2 = \sigma_t^2 \varepsilon_t^2$ y reescribimos la ecuación de volatilidad en (2.36) como

$$\sigma_{t+1}^2(1) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 + \alpha_1 \sigma_t^2 (\varepsilon_t^2 - 1).$$

Cuando $t = h + 1$, la ecuación se convierte en

$$\sigma_{n+2}^2(1) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{n+1}^2 + \alpha_1 \sigma_{n+1}^2 (\varepsilon_{n+1}^2 - 1).$$

Dado que $\mathbb{E}(\varepsilon_{n+1}^2 - 1 | \mathcal{F}_n) = 0$, las predicción para la volatilidad en 2 pasos para un conjunto de observaciones de tamaño n satisface la ecuación

$$\sigma_n^2(2) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_n^2(1).$$

En general, se tiene que

$$\sigma_n^2(k) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_n^2(k-1), \quad k > 1. \quad (2.37)$$

Este resultado es exactamente el mismo que el de un modelo ARMA(1, 1) con un polinomio AR, $1 - (\alpha_1 + \beta_1)B$. Al sustituir repetidamente en (2.36), obtenemos que la predicción en el k -ésimo paso puede ser escrita como

$$\sigma_n^2(k) = \frac{\alpha[1 - (\alpha_1 + \beta_1)^{k-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1} \sigma_n^2(1).$$

Por tanto,

$$\sigma_n^2(k) \rightarrow \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \quad k \rightarrow \infty$$

asumiendo que $\alpha_1 + \beta_1 < 1$. Consecuentemente, la predicción de la volatilidad para varios pasos secuenciales de un modelo GARCH(1,1) converge a la varianza (no condicional) de X_t conforme el horizonte de la predicción tiende a finito asumiendo que $\text{Var}(X_t)$ existe.

El modelo GARCH posee la misma debilidad que el modelo ARCH. En ocasiones, responde de la misma manera a movimientos positivos y negativos. Además, estudios empíricos recientes de series de tiempo financieras de alta frecuencia indican que el compartamiento de las colas de modelos GARCH se mantiene muy pequeño incluso cuando los errores poseen una distribución *t-Student*.

Estimación de parámetros para el modelo GARCH(1,1).

Sea $\Theta = (\alpha_0, \alpha_1, \beta_1)$; la verosimilitud para un conjunto de observaciones provenientes de un procesos GARCH(1, 1) tiene la siguiente forma

$$L(x_1, \dots, x_n | \Theta) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2}x_i^2} \quad (2.38)$$

donde $\sigma_i = \sqrt{\alpha_0 + \alpha_1 x_{i-1}^2 + \beta_1 \sigma_{i-1}^2}$. La optimización de la versosimilitud requiere el uso de un algoritmo numérico para encontrar los estimadores $\hat{\alpha}_0$, $\hat{\alpha}_1$ y $\hat{\beta}_1$ que maximizan a (2.38). Evidentemente para iniciar la implementación numérica se requieren estimadores iniciales para $\hat{\alpha}_0$, $\hat{\alpha}_1$ y $\hat{\beta}_1$, los cuales pueden ser obtenidos mediante las siguientes ecuaciones a partir de la estimación de momentos muestrales de los datos

$$\hat{\sigma}_x^2 = \text{Var}(X_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \quad (2.39)$$

$$\hat{\rho}_1^{x^2} = \frac{\alpha_1(1 - \beta_1^2 - \alpha_1\beta_1)}{1 - \beta_1^2 - 2\alpha_1\beta_1} \quad (2.40)$$

$$\hat{u}_{x^2} = \frac{\hat{\rho}_2^{x^2}}{\hat{\rho}_1^{x^2}} = \alpha_1 + \beta_1 \quad (2.41)$$

$$\widehat{K}(X_t) = \frac{3\{1 - (\alpha_1 - \beta_1)^2\}}{1 - 3\alpha_1^2 - \beta_1^2 - 2\alpha_1\beta_1} \quad (2.42)$$

Si resolvemos las ecuaciones (2.39),(2.40),(2.41) y (2.42) para α_0 , α_1 y β_1 obtenemos las siguientes relaciones

$$\hat{\alpha}_1^2 = \frac{1(\widehat{K}(X_t) - 3)(1 - (\hat{u}_{x^2})^2)}{2\widehat{K}(X_t)} \quad (2.43)$$

$$\hat{\beta}_1 = (\hat{u}_{x^2}) - \hat{\alpha}_1 \quad (2.44)$$

$$\hat{\alpha}_0 = \hat{\sigma}_x^2(1 - \alpha_1 - \beta_1) \quad (2.45)$$

De esta manera (2.43), (2.44) y (2.45) pueden ser usados en (2.38) como valores iniciales para obtener posteriormente los estimadores de máxima verosimilitud $\hat{\alpha}_0$, $\hat{\alpha}_1$ y $\hat{\beta}_1$ para el modelo GARCH(1,1).

2.3.3. Modelo ARMA-GARCH

Una extensión reciente a los modelos ARCH y GARCH ha sido el modelo ARMA-GARCH, el cual contempla un proceso de series de tiempo tanto para la media como para el error, es decir, retomando las ecuaciones (2.34) y (2.10) tenemos que

$$Y_t = \sum_{k=1}^m \phi_k Y_{t-k} + X_t + \sum_{j=1}^n \theta_j X_{t-j}, \quad (2.46)$$

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (2.47)$$

donde nuevamente ε_t es i.i.d. de acuerdo a una cierta distribución $\mathcal{D}_\vartheta(0, 1)$. Un proceso estocástico definidos por las ecuaciones (2.46) y (2.47) recibe el nombre de *proceso* ARMA(m, n)-GARCH(p, q). La estimación de parámetros debe hacerse de manera numérica y los estimadores iniciales pueden

obtenerse al estimar la parte ARMA y GARCH del modelo por separado como si cada uno fuera un modelo independiente del otro. Actualmente en algunas paqueterías especializadas para realizar análisis estadísticos, como MATLAB y R, se cuenta con librerías especializadas para realizar la estimación de parámetros vía máxima verosimilitud en este tipo de modelos. En el siguiente capítulo veremos un ejemplo de cómo instrumentar esta estimación de parámetros en MATLAB para el caso particular del modelo AR-ARCH.

Capítulo 3

Series de tiempo vía variables latentes

Pitt y Walker (2005), han propuesto e instrumentado novedosos modelos de series de tiempo $\{Y_t\}_t$ basándose en la estructura de densidades condicionales $f_{X_t|Y_t=y}$ y $f_{Y_{t+1}|X_t=x}$ que surgen de considerar a un proceso bivariado (Y_t, X_t) estrictamente estacionario.

En esta tesis, propondremos la construcción de un modelo de series de tiempo similar al de Pitt, Walker (2005) pero con la variante de que la media condicional de la serie sigue un proceso AR(1). Este sería un caso de los modelos conocidos en la literatura como ARMA-GARCH (véase Würtz, Chabbi y Luksan, 2006); estos modelos tiene una estructura ARMA en la media condicional, pero las innovaciones o errores son condicionalmente heterocedásticos. En particular se propone que el modelo condicionalmente heterocedástico para las innovaciones sea como los modelos propuestos por Pitt y Walker (2005).

Comenzaremos por estudiar el proceso de innovación $\{Y_t\}_t$ condicionalmente heterocedástico como en Pitt y Walker (2005).

3.1. El modelo de Pitt y Walker

Consideremos una sucesión de variables aleatorias

$$Y_1 \rightarrow W_1 \rightarrow Y_2 \rightarrow W_2 \rightarrow \cdots \rightarrow W_{n-1} \rightarrow Y_n \quad (3.1)$$

donde la variable Y_1 tiene distribución de probabilidades con densidad dada por $f_{Y_1}(y)$. A continuación, condicionalmente al evento $Y_1 = y_1$, W_1 tiene distribución de probabilidades con densidad correspondiente $f_{W_1|Y_1=y}(w|y_1)$. Notemos que lo anterior especifica las relaciones de dependencia estocástica entre W_1 y Y_1 , ya que conocemos la densidad conjunta

$$f_{Y_1, W_1}(y, w) = f_{Y_1}(y) \cdot f_{W_1|Y_1=y}(w|y), \quad (3.2)$$

que a su vez nos permite conocer $f_{Y_1|W_1}(y|w)$. Nuestro objetivo es que el proceso que se describe en (3.1) sea estrictamente estacionario; una forma de hacer esto sería si la densidad conjunta $f_{Y_t, W_t}(y, w)$ es la misma en cada $t = 1, 2, \dots$. Lo anterior implica que las densidades condicionales $f_{W_t|Y_t}(w|y)$ y $f_{Y_{t+1}|W_t}(y|w)$, así como las densidades marginales $f_{Y_t}(y)$ y $f_{W_t}(w)$, no cambian con t .

Además podemos establecer la transición de W_1 a Y_2 usando $f_{Y_2|W_2}(y|w)$, es decir $f_{Y_2|W_1}(y|w_1) \equiv f_{Y_2|W_2}(y|w_1)$, que de acuerdo a lo que hemos discutido es fija y conocida.

En resumen, el proceso (3.1) evoluciona como sigue:

Comenzamos con una variable aleatoria Y_1 con densidad $f_{Y_t}(y)$. Condicionalmente a $Y_t = y_1$ la variable W_1 tiene densidad $f_{W_t|Y_t}(w|y_1)$, condicional a $W_1 = w_1$ la variable aleatoria Y_2 tiene densidad $f_{Y_{t+1}|W_t}(y|w_1) (= f_{Y_t|W_t}(y|w_1))$ y proseguimos en la misma forma para $t = 3, 4, 5, \dots$

La densidad $f_{Y_t|W_t}(y|w)$ es conocida y no cambia con t . Para que ésto suceda especificamos las densidades $f_{Y_t}(y) (= f_Y(y))$ y $f_{W_t|Y_t}(w|y) (= f_{W|Y}(w|y))$.

3.2. Observaciones y variables latentes

Pitt, Chatfield, Walker (2002) y Pitt, Walker (2005) proponen modelos ($f_{W_t|Y_t}$ y $f_{Y_{t+1}|W_t}$) usando el esquema (3.1) donde el proceso $\{Y_1, Y_2, \dots\}$ corresponde a las observaciones de algún fenómeno en la naturaleza y el proceso $\{W_1, W_2, \dots\}$ corresponde a variables latentes que se introducen con el fin de especificar la dependencia estocástica entre Y_{t+1} y Y_t . Dadas las densidades $f_{W_t|Y_t}$ y $f_{Y_{t+1}|W_t}$ tenemos que

$$f_{Y_{t+1}|Y_t}(y|y_t) = \int_{-\infty}^{\infty} f_{Y_{t+1}|W_t}(y|w) f_{W_t|Y_t}(w|y_t) dw \quad (3.3)$$

en el caso de variables $\{W_1, W_2, \dots\}$ absolutamente continuas, o bien

$$f_{Y_{t+1}|Y_t}(y|y_t) = \sum_w f_{Y_{t+1}|W_t}(y|w) f_{W_t|Y_t}(w|y_t)$$

cuando $\{W_1, W_2, \dots\}$ son variables discretas. Por construcción, el proceso $\{Y_1, W_1, Y_2, W_2, \dots\}$ tiene la propiedad de Markov, ya que la densidad de W_t condicional al pasado del proceso sólo depende de Y_t (la observación inmediata anterior), asimismo la densidad de Y_{t+1} condicional al pasado sólo depende de W_t .

Un ejercicio tradicional en procesos de Markov es el verificar que si $\{X_1, X_2, X_3, \dots\}$ es de Markov entonces el proceso $\{X_1, X_3, X_5, \dots\}$ tiene la propiedad de Markov¹. En nuestro contexto lo anterior implica que $\{Y_1, Y_2, \dots\}$ es de Markov. En efecto, por ejemplo para $n = 3$ y asumiendo la existencia de todas las densidades condicionales, tenemos

$$\begin{aligned} f_{Y_3|Y_2Y_1}(y_3|y_2, y_1) &= \frac{f_{Y_3Y_2Y_1}(y_3, y_2, y_1)}{f_{Y_2Y_1}(y_2, y_1)} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f_{Y_3W_2Y_2W_1Y_1}(y_3, w_2, y_2, w_1, y_1)}{f_{Y_2|Y_1}(y_2|y_1) f_{Y_1}(y_1)} dw_2 dw_1 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_{Y_3W_2Y_2W_1Y_1}(y_3, w_2, y_2, w_1, y_1) dw_2 dw_1 \quad (3.4) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_{Y_3|W_2Y_2W_1Y_1}(y_3|w_2, y_2, w_1, y_1) \times f_{W_2|Y_2W_1Y_1}(w_2|y_2, w_1, y_1) \\ &\times f_{Y_2|W_1Y_1}(y_2|w_1, y_1) \times f_{W_1|Y_1}(w_1|y_1) f_{Y_1}(y_1) dw_2 dw_1 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f_{Y_3|W_2}(y_3|w_2) \times f_{W_2|Y_2}(w_2|y_2) \times f_{Y_2|W_1}(y_2|w_1) \\ &\times f_{W_1|Y_1}(w_1|y_1) f_{Y_1}(y_1) dw_2 dw_1 \end{aligned}$$

¹Véase por ejemplo: *An Introduction to Stochastic Processes*, D. Kannan (1979). Sección 3.1.

Para la última igualdad usamos el hecho de que el proceso $\{Y_1, W_1, Y_2, W_2, \dots\}$ es de Markov y el hecho de que $f_{Y_1}(y_1)$ no depende de w_1 y w_2 . Así entonces, de (3.3) y de (3.1)

$$\begin{aligned} f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1) &= \frac{f_{Y_3|Y_2}(y_3|y_2)f_{Y_2|Y_1}(y_2|y_1)f_{Y_1}(y_1)}{f_{Y_2|Y_1}(y_2|y_1)f_{Y_1}(y_1)} \\ &= f_{Y_3|Y_2}(y_3|y_2). \end{aligned}$$

3.3. Un proceso ARCH con variables latentes

Para construir un modelo tipo ARCH estrictamente estacionario, Pitt y Walker (2005) proponen un esquema como en (3.1) donde $f_{W_t|Y_t}(w|y_t) = \mathcal{I}g(w|\frac{\nu+1}{2}, \frac{\nu\beta^2+y_t^2}{2})$ y $f_{Y_t}(y) = \mathcal{S}t(y|0, \frac{1}{\beta^2}, \nu)$, donde $\mathcal{I}g(w|a, b)$ y $\mathcal{S}t(y|0, \mu, \lambda, \alpha)$ denotan las densidades de probabilidad *Gamma Inversa* con parámetros a y b y *t-Student* con parámetros de media μ , parámetro de escala λ y grados de libertad α .

Es fácil ver que dadas las distribuciones que se tienen para Y_t y $W_t|Y_t$, $W_t \sim \mathcal{I}g(\frac{\nu}{2}, \frac{\nu\beta^2}{2})$ y $Y_t|W_t \sim N(0, w)$. Luego la transición para el proceso sería $Y_{t+1}|Y_t \sim \mathcal{S}t(y_{t+1}|0, \lambda = \frac{\nu+1}{\nu\beta^2+(y_t)^2}, \alpha = \nu+1)$. Véase Pitt y Walker (2005) ó el Apéndice B para verificar la anterior.

Teniendo definida la densidad de transición para el proceso $\{Y_t\}_t$ podemos calcular momentos de nuestro interés como su varianza condicional. Tenemos que si $X \sim \mathcal{S}t(x|\mu, \lambda, \alpha)$, $Var(x) = \lambda^{-1} \frac{\alpha}{\alpha-2}$ (ver Bernardo y Smith (1994)), entonces

$$\begin{aligned} Var(y_t|y_{t-1}) &= \frac{\nu\beta^2 + y_{t-1}^2}{\nu + 1} \left(\frac{\nu + 1}{\nu - 1} \right) \\ &= \frac{\nu\beta^2 + y_{t-1}^2}{\nu - 1} \\ \sigma(y_t|y_{t-1}) &= \sqrt{Var(y_t|y_{t-1})} \\ &= \sqrt{\frac{\nu\beta^2}{\nu - 1} + \left(\frac{1}{\nu - 1} \right) y_{t-1}^2}. \end{aligned}$$

De esta manera queda entonces definida la estructura de la varianza condicional de $\{Y_t\}_t$.

Notemos que para el proceso $\{Y_t\}_t$ tenemos que $\mathbb{E}[Y_t|Y_{t-1}] = 0$; ésto se sigue de que $f_{Y_t|Y_{t-1}}$ es una densidad t de Student con media 0. Por otra parte si k es un entero $k \geq 1$

$$\begin{aligned} \text{Cov}[Y_t, Y_{t-k}] &= \mathbb{E}[Y_t \cdot Y_{t-k}] - \mathbb{E}[Y_t]\mathbb{E}[Y_{t-k}] \\ &= \mathbb{E}[Y_t \cdot Y_{t-k}], \\ \text{pues } \mathbb{E}[Y_t] &= \mathbb{E}[\mathbb{E}[Y_t|Y_{t-1}]] = \mathbb{E}[0] = 0, \forall t. \end{aligned}$$

Pero

$$\begin{aligned} \mathbb{E}[Y_t \cdot Y_{t-k}] &= \mathbb{E}[\mathbb{E}[Y_t \cdot Y_{t-k}|Y_{t-1}, Y_{t-2}, \dots]] \\ &= \mathbb{E}[Y_{t-k} \cdot \mathbb{E}[Y_t|Y_{t-1}, Y_{t-2}, \dots]] \\ &= \mathbb{E}[Y_{t-k} \cdot \mathbb{E}[Y_t|Y_{t-1}]], \end{aligned}$$

ya que $\{Y_t\}_t$ es de Markov. Por tanto, $\mathbb{E}[Y_t \cdot Y_{t-k}] = \mathbb{E}[Y_t \cdot 0] = \mathbb{E}[0] = 0$. Entonces $\text{Cov}[Y_t, Y_{t-k}] = 0, \forall k \geq 1$ y como $\{Y_t\}_t$ es estrictamente estacionario $\text{Cov}[Y_t, Y_{t+k}] = 0, \forall k$ entero y $k \neq 0$.

Notemos que la estructura de $\{Y_t\}_t$ es igual a la del modelo ARCH de Engle (1982) pues si hacemos $\alpha_0 = \frac{\nu\beta^2}{\nu-1}$ y $\alpha_1 = (\frac{1}{\nu-1})$ obtenemos

$$\sigma(y_t|y_{t-1}) = \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}.$$

En teoría de series de tiempo, decimos que éste es un modelo para la *volatilidad* de datos heterocedásticos. Sabemos que por definición $\text{Var}(y_t|y_{t-1}) = \mathbb{E}(y_t^2|y_{t-1}) - \mathbb{E}^2(y_t|y_{t-1})$ y por otra parte sabemos que las $\mathbb{E}(y_t|y_{t-1}) = 0$, pues al igual que en un modelo ARCH usual las Y 's no están correlacionadas pero sus cuadrados sí. Usando este hecho concluimos que

$$\mathbb{E}(y_t^2|y_{t-1}) = \sqrt{\frac{\nu\beta^2}{\nu-1} + \left(\frac{1}{\nu-1}\right)y_{t-1}^2}.$$

3.4. Construcción del modelo AR-ARCH

El proceso $\{Y_t\}_t$ jugará el papel de las innovaciones o ruido en un modelo autorregresivo de orden 1 para las observaciones $\{Z_t\}_t$.

Consideremos la ecuación que define a un procesos autorregresivo de orden 1

$$Z_t = \phi Z_{t-1} + Y_t, \quad (3.5)$$

donde $|\phi| < 1$. Podemos pensar en (3.5) como una transformación entre dos vectores aleatorios

$$T(Y_1, \dots, Y_n) = (Z_1, \dots, Z_n),$$

donde dado un valor inicial constante $z_0 = \xi_0$ para el proceso $\{Z_t\}_t$, cada entrada del vector (Z_1, \dots, Z_n) está definida por (3.5).

La transformación inversa

$$T^{-1}(Z_1, \dots, Z_n) = (Y_1, \dots, Y_n)$$

queda definida por

$$Y_t = Z_t - \phi Z_{t-1}. \quad (3.6)$$

Si conocemos la densidad conjunta de (Y_1, \dots, Y_n) entonces el teorema de cambio de variable nos permitirá conocer la densidad conjunta de (Z_1, \dots, Z_n) . Con este fin, el Jacobiano de (3.6) es

$$|J_{T^{-1}}| = \begin{vmatrix} \frac{\partial Y_1}{\partial Z_1} & \frac{\partial Y_1}{\partial Z_2} & \cdots & \frac{\partial Y_1}{\partial Z_n} \\ \frac{\partial Y_2}{\partial Z_1} & \frac{\partial Y_2}{\partial Z_2} & \cdots & \frac{\partial Y_2}{\partial Z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Y_n}{\partial Z_1} & \frac{\partial Y_n}{\partial Z_2} & \cdots & \frac{\partial Y_n}{\partial Z_n} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi & 1 \end{vmatrix} = 1.$$

Así entonces la función de verosimilitud para las observaciones ξ_1, \dots, ξ_n del proceso $\{Z_n\}_n$ está dada por

$$\begin{aligned}
f_{Z_1, \dots, Z_n}(\xi_1, \dots, \xi_n) &= f_{Y_1, \dots, Y_n}(\xi_1 - \phi\xi_0, \dots, \xi_n - \phi\xi_{n-1}) \times |J_{T-1}| \\
&= f_{Y_1}(\xi_1 - \phi\xi_0) \times \prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(\xi_{t+1} - \phi\xi_t | \xi_t - \phi\xi_{t-1})
\end{aligned}$$

donde

$$f_{Y_1}(\xi_1 - \phi\xi_0) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi} \right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{(\xi_1 - \phi\xi_0)^2}{\nu\beta^2} \right]^{\frac{\nu+1}{2}}}$$

la cual es una distribución $\mathcal{St}(\xi_1 | \phi\xi_0, \lambda, \nu)$ con $\lambda = \frac{1}{\beta^2}$.

$$\begin{aligned}
&f_{Y_{t+1}|Y_t}(\xi_{t+1} - \phi\xi_t | \xi_t - \phi\xi_{t-1}) = \\
&= \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})} \left(\frac{1}{(\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2)\pi} \right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{(\xi_{t+1} - \phi\xi_t)^2}{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2} \right]^{\frac{\nu+2}{2}}},
\end{aligned}$$

es decir que la densidad de transición para Y es una $\mathcal{St}(\xi_{t+1} | \phi\xi_t, \lambda, \nu + 1)$ con $\lambda = \frac{\nu+1}{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}$.

En el diagrama de abajo podemos ver el esquema de interrelación entre los datos observables que son las Y 's, las variables latentes que son las W 's y el conjunto de datos completos que son las X 's.

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \dots \rightarrow X_{2(n-1)} \rightarrow X_{2n-1}$$

$$Y_1 \rightarrow W_1 \rightarrow Y_2 \rightarrow W_2 \rightarrow \dots \rightarrow W_{n-1} \rightarrow Y_n$$

Usando los supuestos del artículo de Pitt y Walker (2005), partimos del hecho de que $f_{W_t|Y_t}(w | \xi_t - \phi\xi_{t-1}) = \mathcal{I}g(\frac{\nu+1}{2}, \frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2})$ y $f_{Y_t}(\xi_t - \phi\xi_{t-1}) = \mathcal{St}(\mu = \phi\xi_{t-1}, \lambda = \frac{1}{\beta^2}, \nu)$.

3.5. Usando el algoritmo EM

Una vez que hemos definido perfectamente nuestro modelo, lo que resta es proceder a la estimación de sus parámetros y, como señalamos anteriormente, debido a que estamos considerando un conjunto de variables latentes W lo ideal sería la implementación del algoritmo EM. De esta manera, la expresión de la verosimilitud aumentada considerando el conjunto de datos completos sería la siguiente

$$f_{Z_1, \dots, Z_n, W_1, \dots, W_{n-1}} | \nu, \beta, \phi(\xi_1, \dots, \xi_n, w_1, \dots, w_{n-1}) = \prod_{t=1}^{n-1} \left\{ f_{Y_{t+1}|W_t}(\xi_{t+1} - \phi\xi_t | w_t) \right. \\ \left. \times f_{W_t|Y_t}(w_t | \xi_t - \phi\xi_{t-1}) \times f_{Y_1}(\xi_1 - \phi\xi_0) \right\}.$$

Consideremos $\pi(\nu, \beta, \phi) \propto 1$, así

$$\begin{aligned} \pi(\nu, \beta, \phi | z_1, \dots, z_n, w_1, \dots, w_{n-1}) &\propto f_{Z_1, \dots, Z_n, W_1, \dots, W_{n-1}} | \nu, \beta, \phi \times \pi(\nu, \beta, \phi) \\ &= f_{Z_1, \dots, Z_n, W_1, \dots, W_{n-1}} | \nu, \beta, \phi. \end{aligned}$$

Como se dijo en la sección 1.2.1 si la distribución *a priori* es proporcional a una constante, la distribución *a posteriori* es proporcional entonces a la verosimilitud. Luego en la sección 1.4 se mencionó que el objetivo del algoritmo EM es encontrar la moda de la distribución *a posteriori* y, recordando que si la distribución *a posteriori* es proporcional a la verosimilitud, al encontrar su moda estaríamos encontrando el estimador o los estimadores máximo verosímiles. Esto es lo que se pretende al usar el algoritmo EM, encontrar a $\hat{\nu}$, $\hat{\beta}$ y $\hat{\phi}$.

Ahora lo que nos interesa es construir la distribución predictiva para los datos aumentados y ésto será obtener lo siguiente

$$\begin{aligned} \frac{f_{Z_1, \dots, Z_n, W_1, \dots, W_{n-1}} | \nu, \beta, \phi}{f_{Z_1, \dots, Z_n} | \nu, \beta, \phi} &= \frac{\prod_{t=1}^{n-1} \left\{ f_{Y_{t+1}|W_t}(\xi_{t+1} - \phi\xi_t | w_t) f_{W_t|Y_t}(w_t | \xi_t - \phi\xi_{t-1}) \right\}}{\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(\xi_{t+1} - \phi\xi_t | \xi_t - \phi\xi_{t-1})} \\ f_{W_1, \dots, W_{n-1}} | Z_1, \dots, Z_n, \nu, \beta, \phi &= \prod_{t=1}^{n-1} \left\{ \frac{f_{Y_{t+1}|W_t}(\xi_{t+1} - \phi\xi_t | w_t) f_{W_t|Y_t}(w_t | \xi_t - \phi\xi_{t-1})}{f_{Y_{t+1}|Y_t}(\xi_{t+1} - \phi\xi_t | \xi_t - \phi\xi_{t-1})} \right\}. \end{aligned}$$

Tomado logaritmo en la densidad *a posteriori* se obtiene lo siguiente

$$\begin{aligned} \log[\pi(\nu, \beta, \phi|z_1, \dots, z_n, w_1, \dots, w_{n-1})] &\triangleq \sum_{t=1}^{n-1} \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(w_t) \right. \\ &- \frac{1}{2w_t} (\xi_{t+1} - \phi\xi_t)^2 + \left(\frac{\nu+1}{2}\right) \log\left(\frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2}\right) - \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) \\ &\left. - \left(\frac{\nu+1}{2} + 1\right) \log(w_t) - \left(\frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2}\right) w_t^{-1} \right] + \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) \\ &- \frac{1}{2} \log(\nu\beta^2\pi) - \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{(\xi_1 - \phi\xi_0)^2}{\nu\beta^2}\right). \end{aligned}$$

3.5.1. Paso - E

Recordemos que en la etapa *E* se busca tomar la esperanza sobre la variable latente en la expresión anterior y así obtener $Q(\Theta, \Theta^{(k)})$, donde $\Theta' = (\nu, \beta, \phi)$ y (k) denota el valor actual de Θ .

$$\begin{aligned} Q(\Theta, \Theta^{(k)}) &= \sum_{t=1}^{n-1} \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \mathbb{E}[\log(w_t)|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] - \frac{(\xi_{t+1} - \phi\xi_t)^2}{2} \right. \\ &\times \mathbb{E}[w_t^{-1}|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] + \left(\frac{\nu+1}{2}\right) \log\left(\frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2}\right) - \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \\ &\left. \left(\frac{\nu+3}{2}\right) \right. \\ &\times \mathbb{E}[\log(w_t)|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] - \left.\left(\frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2}\right) \mathbb{E}[w_t^{-1}|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] \right] \\ &+ \log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right) - \log\left(\Gamma\left(\frac{\nu}{2}\right)\right) - \frac{1}{2} \log(\nu\beta^2\pi) - \left(\frac{\nu+1}{2}\right) \log\left(1 + \frac{(\xi_1 - \phi\xi_0)^2}{\nu\beta^2}\right). \end{aligned}$$

Renombrando las expresiones donde tomamos esperanzas

$$\begin{aligned} \eta_t^{(k)} &\equiv \mathbb{E}[\log(w_t)|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] \\ \mu_t^{(k)} &\equiv \mathbb{E}[w_t^{-1}|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}]. \end{aligned}$$

Pasaremos ahora al cálculo de estas dos expresiones.

$$\mathbb{E}[\log(w_t) | \xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] = \frac{1}{f_{Y_{t+1}|Y_t(\xi_{t+1}-\phi^{(k)}\xi_t|\xi_t-\phi^{(k)}\xi_{t-1})}} \times$$

$$\int_0^\infty \log(w) \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} \right)^{\frac{\nu^{(k)}+1}{2}}}{\Gamma\left(\frac{\nu^{(k)}+1}{2}\right)} w^{-(\frac{\nu^{(k)}+2}{2}+1)}$$

$$\times e^{-\frac{\left(\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{2} \right)}{w}} dw$$

$$= \frac{(\pi[\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2])^{\frac{1}{2}}}{\Gamma\left(\frac{\nu^{(k)}+2}{2}\right)} \frac{\left[1 + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2} \right]^{\frac{\nu^{(k)}+2}{2}}}{\sqrt{2\pi}}$$

$$\left(\frac{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} \right)^{\frac{\nu^{(k)}+1}{2}} \times \int_0^\infty \log(w) w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw,$$

$$\text{con } a^{(k)} = \frac{\nu^{(k)}+2}{2}, b^{(k)} = \frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{2}.$$

(Ver el cálculo de $\int_0^\infty \log(w) w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw$ en el Apéndice A).

$$\begin{aligned}
\mathbb{E}[\log(w_t)|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] &= \frac{1}{\sqrt{2\pi}} \sqrt{\pi} (\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1}))^{\frac{1}{2}} \\
&\times \frac{\left[\log \left(\frac{\nu^{(k)} (\beta^{(k)})^2}{2} \right) + \frac{(\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)} \xi_t)^2}{2} \right]}{\left(\frac{\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} \right)^{\frac{1}{2}}} \\
&- \frac{\psi \left(\frac{\nu^{(k)} + 2}{2} \right)}{\left(\frac{\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} \right)^{\frac{1}{2}}} \\
&= \log \left(\frac{\nu^{(k)} (\beta^{(k)})^2}{2} \right) + \frac{(\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} \\
&+ \frac{(\xi_{t+1} - \phi^{(k)} \xi_t)^2}{2} - \psi \left(\frac{\nu^{(k)} + 2}{2} \right) \equiv \eta_t^{(k)}
\end{aligned}$$

donde $\psi(\cdot) \equiv \frac{d}{dt} \Gamma(t)$ es conocida como *función digamma*.

$$\begin{aligned}
\mathbb{E}[w_t^{-1}|\xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] &= \frac{1}{f_{Y_{t+1}|Y_t}(\xi_{t+1} - \phi^{(k)} \xi_t | \xi_t - \phi^{(k)} \xi_{t-1})} \times \\
&\int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\nu^{(k)} (\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} \right)^{\frac{\nu^{(k)} + 1}{2}}}{\Gamma\left(\frac{\nu^{(k)} + 1}{2}\right)} w^{-\left(\left(\frac{\nu^{(k)} + 2}{2} + 1\right) + 1\right)} \\
&\times e^{-\frac{\left(\frac{\nu^{(k)} (\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)} \xi_t)^2}{2} \right)}{w}} dw \\
&= \frac{(\pi [\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1})^2])^{\frac{1}{2}}}{\Gamma\left(\frac{\nu^{(k)} + 2}{2}\right)} \frac{\left[1 + \frac{(\xi_{t+1} - \phi^{(k)} \xi_t)^2}{\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1})^2} \right]^{\frac{\nu^{(k)} + 2}{2}}}{\sqrt{2\pi}} \\
&\left(\frac{\nu^{(k)} (\beta^{(k)})^2 + (\xi_t - \phi^{(k)} \xi_{t-1})^2}{2} \right)^{\frac{\nu^{(k)} + 1}{2}} \times \int_0^\infty w^{-(\alpha^{(k)} + 1)} e^{-\frac{b^{(k)}}{w}} dw,
\end{aligned}$$

$$\text{con } \alpha^{(k)} = \frac{\nu^{(k)}+2}{2} + 1, b^{(k)} = \frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{2}.$$

(Ver el cálculo de $\int_0^\infty w^{-(\alpha^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw$ en el Apéndice A).

$$\begin{aligned} \mathbb{E}[w_t^{-1} | \xi_{t+1}, \xi_t, \xi_{t-1}, \Theta^{(k)}] &= \frac{(\frac{\nu^{(k)}+2}{2})2}{\left[\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2 \right]} \\ &\times \frac{1}{\left[1 + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2} \right]} \\ &= \frac{\nu^{(k)} + 2}{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2 + (\xi_{t+1} - \phi^{(k)}\xi_t)^2} \\ &\equiv \mu_t^{(k)} \end{aligned}$$

3.5.2. Paso - M

Después de efectuar el paso-E lo siguiente es maximizar $Q(\Theta|\Theta^{(k)})$ y para ésto primero debemos derivar esta función con respecto a cada uno de los parámetros ϕ, β y ν . Las expresiones para las derivadas serían las que se presentan a continuación:

$$\begin{aligned} \frac{\partial Q}{\partial \phi} &= \sum_{t=1}^n \left[\mu_t^{(k)} (\xi_{t+1} - \phi \xi_t) \xi_t - \left(\frac{\nu+1}{2} \right) \left[\frac{2}{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2} \right] (\xi_t - \phi\xi_{t-1}) \xi_{t-1} \right. \\ &\quad \left. + \mu_t^{(k)} (\xi_t - \phi\xi_{t-1}) \xi_{t-1} \right] + \left(\frac{\nu+1}{2} \right) \left(\frac{\nu\beta^2}{\nu\beta^2 + (\xi_1 - \phi\xi_0)^2} \right) \left(\frac{2(\xi_1 - \phi\xi_0)}{\nu\beta^2} \right) \xi_0 \end{aligned}$$

$$\begin{aligned} \frac{\partial Q}{\partial \beta} &= \sum_{t=1}^n \left[\left(\frac{\nu+1}{2} \right) \left(\frac{2}{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2} \right) \nu\beta - \mu_t^{(k)} (\nu\beta) \right] \\ &\quad - \frac{1}{\beta} + \left(\frac{\nu+1}{2} \right) \left(\frac{\nu\beta^2}{\nu\beta^2 + (\xi_1 - \phi\xi_0)^2} \right) \left(\frac{2(\xi_1 - \phi\xi_0)^2}{\nu\beta^3} \right) \end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \nu} &= \sum_{t=1}^n \left[\left(\frac{\nu+1}{2} \right) \left[\frac{\beta^2}{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2} \right] + \frac{1}{2} \log \left(\frac{\nu\beta^2 + (\xi_t - \phi\xi_{t-1})^2}{2} \right) \right. \\
&\quad \left. - \frac{1}{2} \psi \left(\frac{\nu+1}{2} \right) - \frac{1}{2} \eta_t^{(k)} - \frac{\beta^2}{2} \mu_t^{(k)} \right] + \frac{1}{2} \psi \left(\frac{\nu+1}{2} \right) - \frac{1}{2} \psi \left(\frac{\nu}{2} \right) - \frac{1}{2\nu} \\
&\quad - \frac{1}{2} \log \left(1 + \frac{(\xi_1 - \phi\xi_0)^2}{\nu\beta^2} \right) + \left(\frac{\nu+1}{2} \right) \left[\frac{1}{\nu\beta^2 + (\xi_1 - \phi\xi_0)^2} \right] \frac{(\xi_1 - \phi\xi_0)^2}{\nu}.
\end{aligned}$$

Posteriormente se tienen que igualar a cero las 3 ecuaciones simultáneamente y debido a que su solución no tiene una forma analítica cerrada, lo que se hizo para resolver el sistema fue usar el *método de Newton* y así dar soluciones aproximadas. Basta mencionar que el método de Newton se itera hasta que las raíces de la 3 ecuaciones converjan y el algoritmo EM se repite hasta que $\|\Theta^{(k+1)} - \Theta^{(k)}\| < \epsilon$, donde ϵ es la tolerancia deseada.

3.6. Simulación y ajuste del modelo

En esta sección mostraremos un ejemplo en el que se pueden realizar la estimación de parámetros en este tipo de modelos mediante la simulación de una serie de datos. Para evaluar el proceso de estimación de parámetros usaremos varias funciones de Matlab. Algunas de estas funciones fueron creadas para esta tesis y otras se pueden encontrar disponibles en librerías especializadas de Matlab (llamadas *Toolboxes*). Al final del trabajo en el Apéndice B se enlistan aquellas funciones que no forman parte de las librerías de Matlab.

Primeramente se hizo una simulación de observaciones de acuerdo al modelo de Pitt y Walker. Posteriormente se procedió a estimar los parámetros de la serie simulada lo cual se hizo en dos etapas. En la primera se ajustó el modelo como si fuera un proceso autoregresivo ordinario y, ya que sabemos que una primera estimación puede hacerse con una simple regresión, se aplicó el método de mínimos cuadrados para estimar los parámetros del modelo.

A nivel de aproximación podemos considerar a los residuales Y_1, \dots, Y_n como una muestra i.i.d. de una distribución $\mathcal{St}(y|0, \beta^2, \nu)$. Entonces maximizamos la verosimilitud

$$L(\beta, \nu) = \prod_{t=1}^n f_{Y_t}(y|0, \beta^2, \nu),$$

como función de β y ν para obtener estimadores preliminares de estos parámetros. De esta forma obtenemos estimaciones preliminares para ϕ, β y ν .

En la segunda etapa se usan las estimaciones iniciales que se obtuvieron en la etapa anterior para introducirlas en la estimación mediante el algoritmo EM y esto se hace con la función *EM_est.m* (esta función a su vez depende de otra la cual es *newtonQ.m*). Ambas etapas se implementan conjuntamente en la función *fitparams.m*

Una vez que se ha obtenido la estimación final $\hat{\phi}, \hat{\beta}$ y $\hat{\nu}$ de los parámetros, podemos calcular los residuales $\{Y_t\}_t$ del modelo AR(1) para los retornos $\{Z_t\}_t$ y con esta misma información construimos el proceso de volatilidad.

$$\text{Var}(Y_t|Y_{t-1}) = \frac{y_{t-1}^2 + \beta^2\nu}{\nu - 1}$$

Esta parte se hizo con la función *fitmodel.m*. Para ejemplificar este proceso, se hizo una simulación de $\{Z_t\}_t$ mediante la función *AR_ARCH.m*. Las estimaciones de ϕ, β, ν y $\{Y_t\}_t$ obtenidas mediante el método descrito arriba, se comparan con la estimación que hace Matlab con el *Toolbox GARCH*. Esta es una librería especializada para modelos ARMA-GARCH. En la función *ar_archfit.m* se combinan otros comandos de esta librería como son *garchset.m* y *garchfit.m*.

Esto permite comparar el ajuste mediante estas funciones de Matlab y el ajuste propuesto por nosotros que utiliza el algoritmo EM. A manera de resumen, lo que se hizo fue simular 2500 observaciones que siguen el proceso que estamos planteando, es decir, un modelo AR(1) para los retornos $\{Z_t\}_t$ con errores que siguen un proceso ARCH(1) como plantea el artículo de Pitt y Walker (2005).

Para verificar la efectividad de predicción en el modelo, se quitaron las últimas 5 observaciones para ajustar el modelo con las observaciones restantes y luego se realizaron 5 predicciones para compararse con las 5 observaciones reales que fueron removidas del conjunto total de datos. Veamos la instrumentación que se llevó a cabo en Matlab.

```
>> [r vol]=AR_ARCH(2.5,4.5,0.95,2500,500);
>> [beta nu phi]=fitparams(r,5);
```

En esta parte se indica en la función que van a simularse 3000 datos de los cuales las primeras 500 observaciones serán ignoradas (por lo cual se toma 500 como el tamaño del período de calentamiento) quedando así 2500 datos simulados. Los parámetros que se escogieron para la simulación como valores de ϕ , β y ν son 0.95, 2.5 y 4.5 respectivamente. Como resultado se obtendrán un conjunto de datos (r) y un proceso de volatilidad asociado a los datos (vol).

```
El valor de beta es 2.5036478
El valor de nu es 4.4077945
El valor de phi es 0.9485942
```

Se puede observar que las estimaciones finales de los parámetros son muy cercanas a los verdaderos valores.

```
>> [xp sigma resid]=fitmodel(r,vol,beta,nu,phi,5);
```

Dato Real	Predicción	Vol. Real	Predicción
8.748956	9.146412	3.136305	3.166472
4.194797	8.676233	2.833945	3.324136
2.371403	8.230224	3.574783	3.369001
0.304253	7.807143	2.953432	3.382054
0.575444	7.405810	3.011085	3.385875

Arriba se muestran las predicciones comparadas con su valor real tanto para los datos como para la volatilidad. Como es de esperarse, debido a los argumentos señalados en el capítulo 3 en la sección del proceso AR, solo la primera predicción es cercana al verdadero valor mientras que las demás ya son muy alejas; lo anterior se debe a que el modelo es un AR(1) en primera

instancia y en segunda a la naturaleza del modelo al ser la varianza heterocedástica, usualmente hay mucha incertidumbre en tales modelos como para realizar predicciones a más de un paso; comunmente es preferible realizar solo una predicción y una vez que se tengan más observaciones se reajusta el modelo y se realiza la siguiente predicción. Esto de ilustrará en el ejemplo en el que se ajusta un modelo para datos reales.

Ahora haremos el ajuste de parámetros usando las librerías de Matlab.

```
>> x2=r(1:2495);
>> ar_archfit(x2,5)
```

```
beta0 =
    2.5884
```

```
nu0 =
    2.9417
```

```
phi0 =
    0.9481
```

Podemos ver que las estimaciones para ϕ y β son cercanas a los verdaderos valores mientras que la estimación para ν no es nada cercana. Esto se debe a que las librerías de Matlab no está diseñadas para ajustar modelo ARMA-GARCH con un distribución para el error que no sea estandarizada, es decir, en este caso aunque el error del proceso es una *t-Student* reescalada, Matlab ajusta el modelo como si el error fuera una *t-Student* con media cero y varianza uno.

Dato Real	Predicción	Vol. Real	Predicción
8.748956	9.0825	3.136305	2.8632
4.194797	8.5520	2.833945	3.0368
2.371403	8.0491	3.574783	3.1224
0.304253	7.5722	2.953432	3.1656
0.575444	7.1200	3.011085	3.1876

Las predicciones en este caso fueron muy similares a las que se obtuvieron con nuestro modelo, sin embargo, la predicción a un paso para la volatilidad no fue tan cercana como la nuestra.

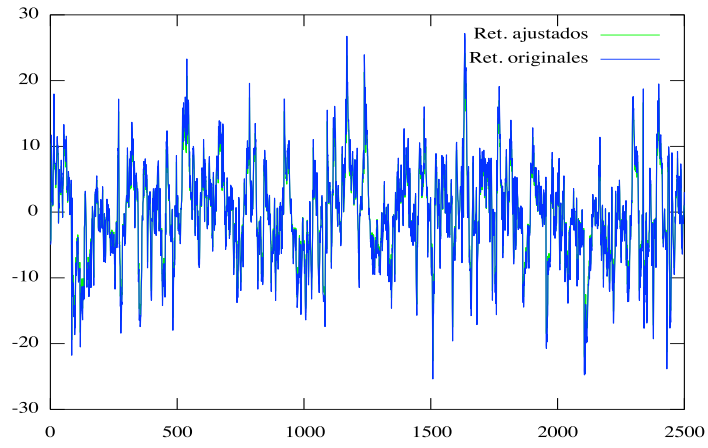


Figura 3.1: Datos estimados y originales

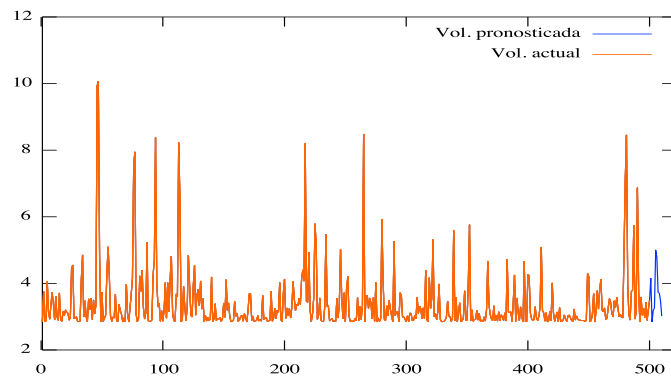


Figura 3.2: Volatilidad estimada y real

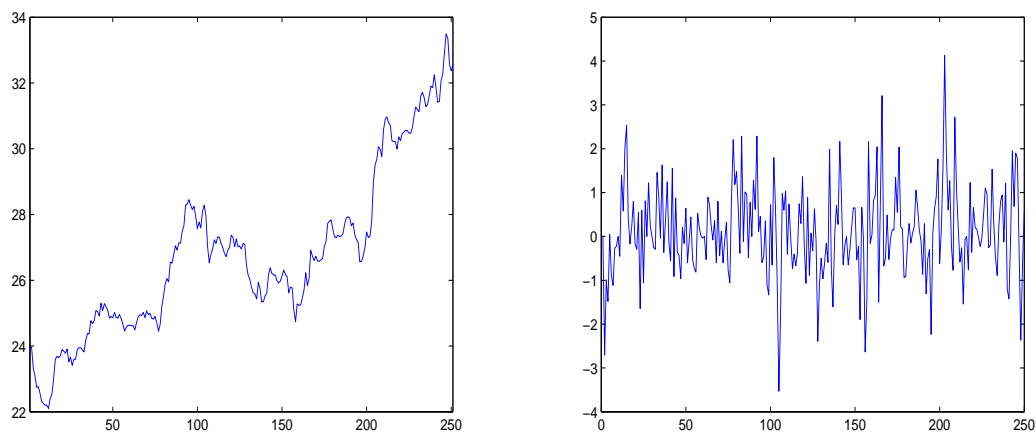


Figura 3.3: Precios y retornos de la serie AT&T

En la figura 3.1 y 3.2 se muestran la serie de datos originales junto con la serie de datos ajustados y la volatilidad real junto con la estimada. En ambos vemos una similitud sobre todo para la volatilidad lo cual es consistente con nuestras estimaciones.

3.7. Aplicación a datos financieros

Veamos ahora la aplicación de nuestro modelo en una serie financiera. La serie que se modeló fueron precios históricos de la acción de la compañía americana AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006, lo cual hace un total de 251 datos. Estos datos pueden descargarse en la página web <http://finance.yahoo.com/>. A continuación se presenta el ajuste de la serie con las funciones programadas en MATLAB.

El análisis exploratorio de una serie de tiempo como sabemos es fundamental y algo que se puede observar a simple vista en la parte derecha de la figura 3.3 son los “picos” repentinos en la serie y ésto nos habla de que no hay homogeneidad en la varianza. Es por ello que cabe la hipótesis de un modelo del tipo ARCH ó GARCH. Pero, por otra parte, al examinar el correlograma

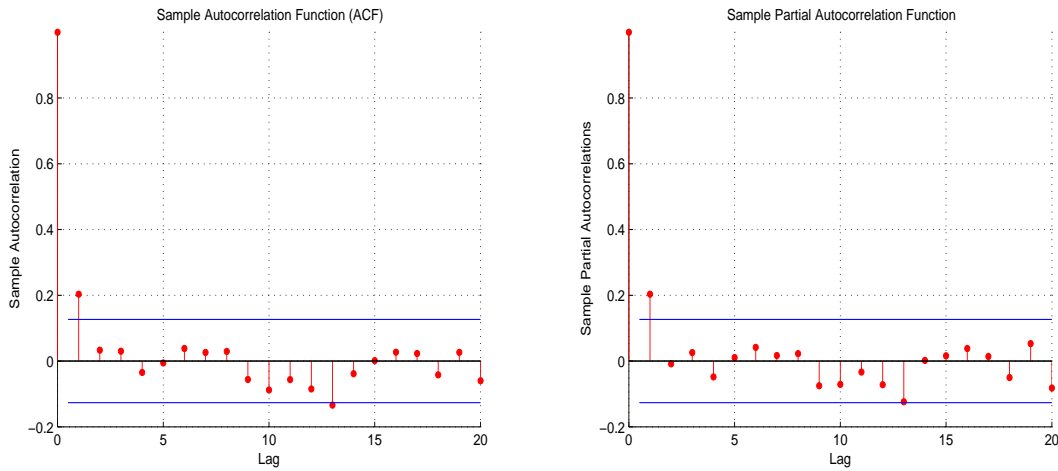


Figura 3.4: Correlograma de los retornos

podemos ver que sí hay autocorrelación; en particular, si nos fijamos en la función de autocorrelación parcial vemos que sólo el primer retraso se sale de las bandas de confianza como se puede apreciar en la figura 3.4, lo cual nos da motivos para sugerir un modelo AR(1).

Veamos qué sucede al ajustar a la serie de datos un modelo AR(1) ordinario, es decir, con errores que tienen distribución Normal.

```
>> att=load('att.txt');
>> ret=price2ret(att);
>> r_att=ret'*100;
>> [datos resid phi0]=AR_OLS(r_att,1,1);
```

Primeramente se lee el archivo con la serie de precios de la acción; posteriormente se calculan los retornos de la serie y se reescalan al multiplicarse por 100; en el libro de Fan y Yao (2003), pág. 171, la serie de retornos financieros se multiplica por 100 en un ejemplo que presentan para evitar trabajar con valores numéricamente pequeños. Por último se aplica el método de mínimos cuadrados para estimar el parámetro del modelo AR(1).

Si observamos la figura 3.5, considerar una distribución normal para los residuales no es la mejor alternativa de modelo ya que, como se puede ob-

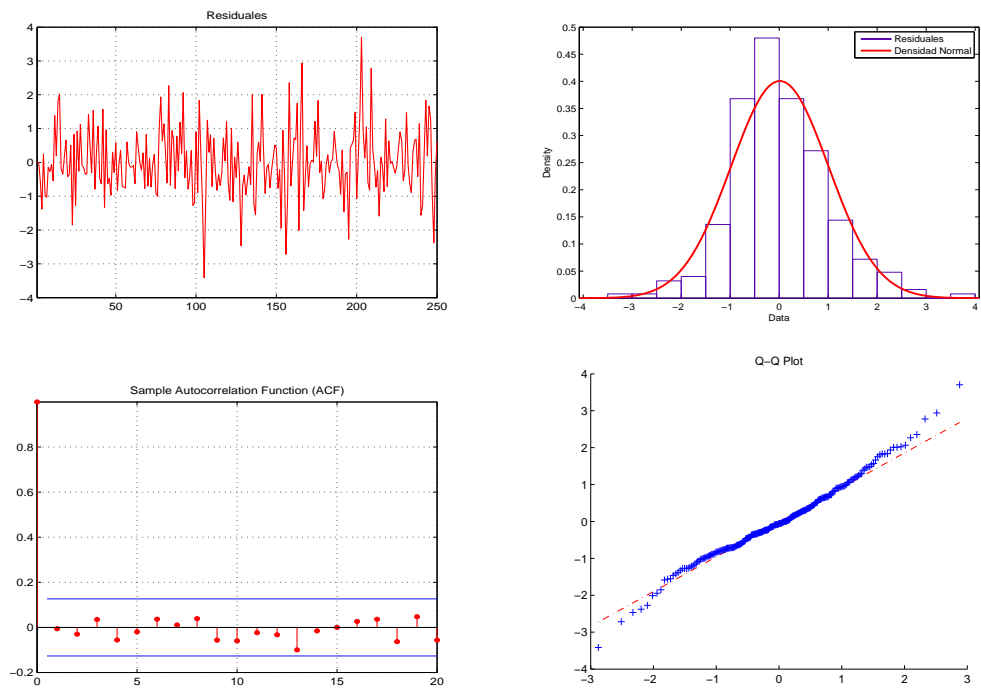


Figura 3.5: Residuales del modelo AR(1) ordinario

servar en el histograma y en el *QQ-plot*, tiene exceso de curtosis. Además se realizó la prueba de Jarque-Bera en donde se plantea la hipótesis nula de que las observaciones siguen una distribución normal contra que no, la cual se rechazó como se muestra en los comandos de Matlab.

```
>> res_analysis(resid)
```

```
H =
```

```
1
```

```
p_value =
```

```
0.0053
```

```
T =
```

```
14.7412
```

Ahora, vamos a contrastar los resultados anteriores con los obtenidos al ajustar los datos con el modelo que se propone en este trabajo. Estos son los parámetros resultantes que se obtuvieron al implementar el algoritmo EM.

```
>> [beta nu phi]=fitparams(r_att);
```

```
beta0      nu0      phi0
0.8443     6.9025     0.2144
```

```
El valor de beta es: 0.89818
```

```
El valor de nu es: 8.61116
```

```
El valor de phi es: 0.17239
```

Una vez que hemos obtenidos las estimaciones para los parámetros del modelo, vamos a generar los residuales con la siguiente línea de código. Cabe mencionar que esta misma función (*fitSeries.m*) se usará más adelante para realizar predicciones con el modelo y en este caso solo deseamos calcular los

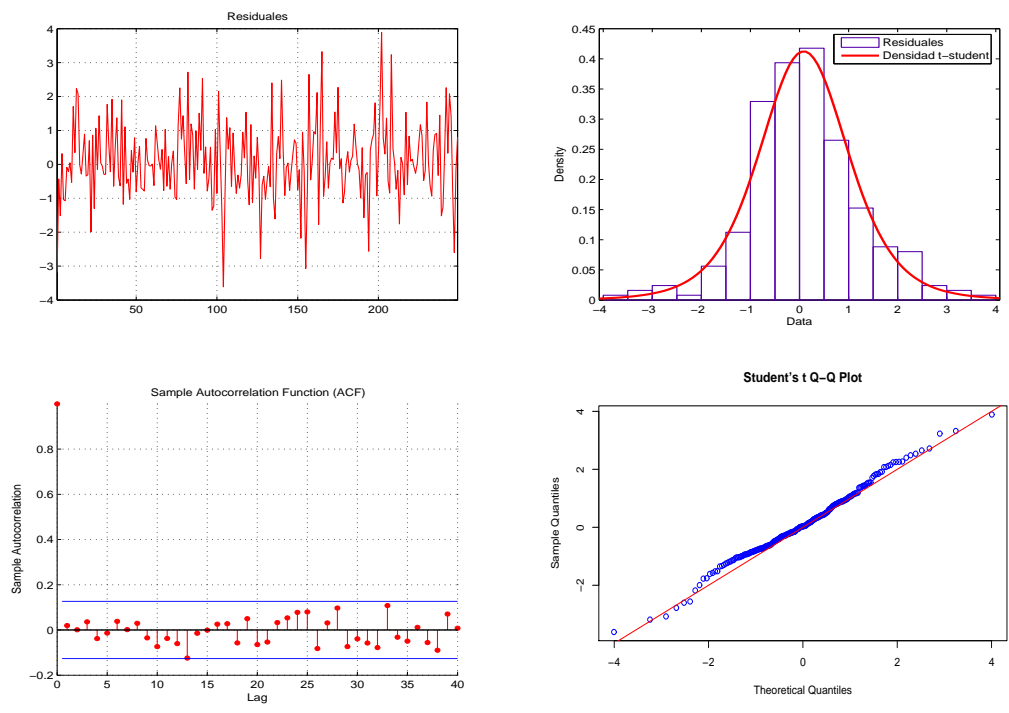


Figura 3.6: Residuales de nuestro modelo

residuales por lo que el último parámetro vale un cero ya que especifica las precesiones que quieren realizarse (ver código completo en el Apéndice B para detalles de la función).

```
>> [sig res pred]=fitSeries(r_att,beta,nu,phi,0);
```

Revisemos así el análisis de residuales del modelo ARCH $\{Y_t\}_t$ para verificar la bondad del ajuste realizado. En la figura 3.6 se puede apreciar, tanto en el histograma como el *QQ-plot*, que la distribución de los residuales se asemeja mucho a una $St(x|0, 1, \nu + 1)$ y esta hipótesis es justificada al aplicar la prueba de Kolmogorov-Smirnov en las cual se constrasta la hipótesis

$$H_0 : r_i \sim St(x|0, 1, \nu + 1) \quad V.S. \quad H_1 : r_i \not\sim St(x|0, 1, \nu + 1), \quad i = 1, \dots, n,$$

donde r_i es el i -ésimo residual y en dicha prueba podemos ver que no se rechaza H_0 .

```
>> res_analysis(res,nu+1)
```

```
H =
```

```
0
```

```
p_value =
```

```
0.2288
```

```
T =
```

```
0.0654
```

Al aplicar la prueba de Kolmogorov-Smirnov el *p-valor* es grande y por tanto no se rechaza la hipótesis de que la distribución para los residuales de la parte ARCH sigan un disitribución *t-Student* estandarizada.

Por otra parte, en el correlograma se aprecia que no hay autocorrelación en los residuales obtenidos pero parece que el retraso 13 se sale un poco de las bandas de confianza. Para eliminar la duda se realizó la prueba de Box-Ljung, la cual contrasta las hipótesis

$$H_0 : \rho_0 = \rho_1 = \dots = \rho_n = 0 \quad V.S. \quad H_1 : \rho_i \neq 0 \quad p.a. \quad i, \quad i = 1, \dots, n.$$

Aplicando esta prueba en el paquete *R* se obtuvieron los siguientes resultados:

```
> Box.test(x, lag = 13, type = "Ljung", fitdf = 0)
```

```
Box-Ljung test
```

```
data: x
X-squared = 8.5925, df = 13, p-value = 0.803
```

Como podemos ver, no se rechaza H_0 con lo cual podemos concluir que nuestro modelo describe los datos en forma adecuada. Lo siguiente será comparar la bondad de ajuste de este modelo contra la de los modelos tradicionales que no consideran las existencia de variables latentes.

El primer modelo a comparar es un MA(1)-ARCH(1) con errores $St(0, 1, \nu)$. Haciendo uso de las librerías *GARCH Toolbox* de Matlab se pueden realizar ajustes de modelos de la forma AMA(r, m)-GARCH(q, p) e incluso otras variantes del mismo. Veamos el ajuste realizado para nuestros datos bajo este modelo.

```
>> [coef,error,llf,e,sigma_y] = arma_garchfit(r_att,0);
```

```
Mean: ARMAX(0,1,0); Variance: GARCH(0,1)
```

```
Conditional Probability Distribution: T
```

```
Number of Model Parameters Estimated: 5
```

Parameter	Value	Standard Error	T Statistic
-----	-----	-----	-----

C	0.094701	0.070457	1.3441
MA(1)	0.15615	0.061527	2.5379
K	1.0148	0.18137	5.5950
ARCH(1)	0.027313	0.10356	0.2637
DoF	6.16	3.0208	2.0392

Log Likelihood Value: -352.813

Arriba se muestran los valores estimados para los parámetros del modelo así como otras estadísticas relevantes.

Vamos a realizar la misma prueba de hipótesis para la distribución de los residuales que realizamos en nuestro modelo.

```
>> dof=garchget(coef,'DoF');
```

```
>> res_analysis(e,dof)
```

H =

0

p_value =

0.1948

T =

0.0677

Como vemos el *p-valor* es grande y no se rechaza H_0 . En la figura 3.7 se muestra una gráfica de los residuales, el histograma, el *QQ-plot* y el correlograma.

```
> Box.test(y1, lag = 1, type = "Ljung", fitdf = 0)
```

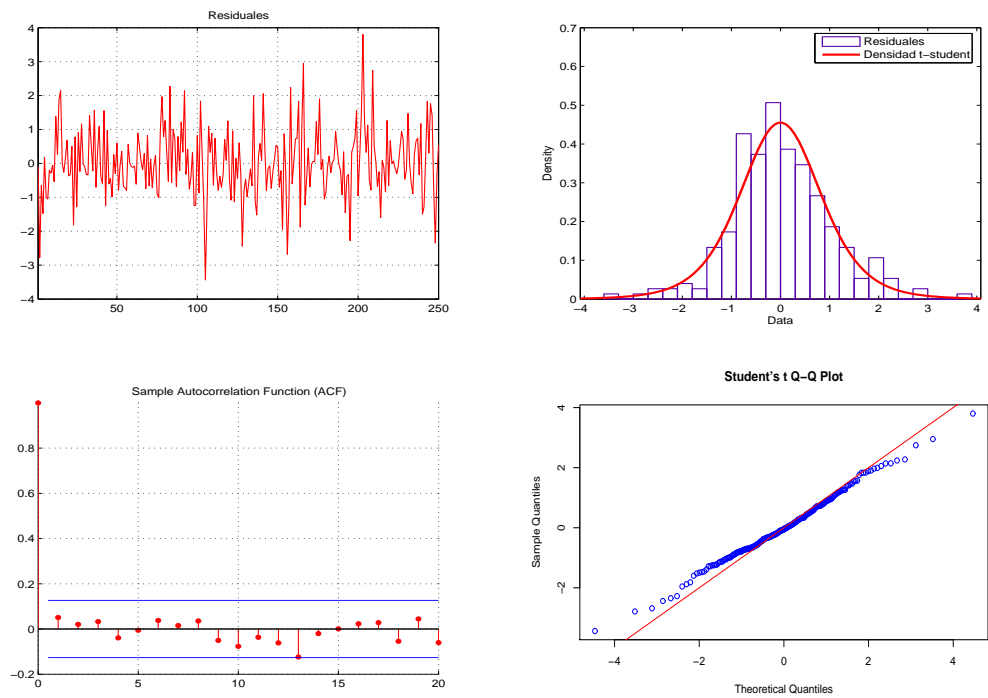


Figura 3.7: Residuales del modelo MA(1) - ARCH(1)

Box-Ljung test

```
data: y1
X-squared = 0.6553, df = 1, p-value = 0.4182
```

En los residuos de este modelo tampoco se rechazó la prueba de Kolmogorov-Smirnov ni la prueba de Box-Ljung, aún así, si observamos el histograma vemos que estos no se ajustan tan bien a una distribución *t-Student* y el *QQ-plot* nos muestra que nuestro modelo produce residuos más cercanos a tener distribución *t-Student*.

Ahora veamos el ajuste con un modelo AR(1)-ARCH(1) ordinario con distribución *t-Student*.

```
>> [coef,error,llf,e,sigma_y] = arma_garchfit(r_att,0);
```

```
Mean: ARMAX(1,0,0); Variance: GARCH(0,1)
```

```
Conditional Probability Distribution: T
```

```
Number of Model Parameters Estimated: 5
```

Parameter	Value	Standard Error	T Statistic
C	0.078282	0.061113	1.2809
AR(1)	0.1652	0.059379	2.7821
K	1.0165	0.18281	5.5601
ARCH(1)	0.02392	0.10326	0.2316
DoF	6.1532	2.994	2.0552

```
Log Likelihood Value: -352.591
```

En la parte de arriba se muestran las estimaciones obtenidas para los parámetros del modelo junto con sus estadísticas y a continuación se muestran los resultados de la prueba de hipótesis para la distribución de los residuales.

```
>> dof=garchget(coef,'DoF');
```

```
>> res_analysis(e,dof)
```

```
H =
```

```
0
```

```
p_value =
```

0.2154

T =

0.0662

Vemos que no podemos rechazar la hipótesis nula. Nuevamente realizaremos la prueba de Box-Ljung para descartar la posibilidad de autocorrelación en los residuales.

```
> Box.test (y2, lag = 1, type="Ljung")
```

Box-Ljung test

data: y2

X-squared = 0.4019, df = 1, p-value = 0.5261

Los residuales de este modelo son muy similares a los del modelo anterior y también presentan las mismas deficiencias: exceso de curtosis respecto a la distribución *t-Student* como se puede observar en la figura 3.8, no se rechazó ni la prueba de Kolmogorov-Smirnov ni la de Box-Ljung.

Quizás estos no son elementos suficientes para exhibir la ventaja de manejar un modelo con variables latentes a uno tradicional pero, sí debemos notar que nuestro modelo posee un parámetro extra, β , el cual es precisamente el resultado de incorporar un proceso latente en el modelo y nos da una distribución más general para el error, una *t-Student* reescalada, mientras que las librerías de Matlab solo consideran la distribución *t-Student* con media cero y varianza unitaria que no es tan flexible y no alcanza a capturar características importantes presentes en los datos como son la asimetría, que como vimos en el capítulo 3, es una característica representativa de los modelos ARCH y GARCH. Por tanto, el hecho de incluir en los datos una distribución más general, al considerar variables latentes, nos da más argumentos para sustentar que nuestro modelo nos ayuda a describir mejor los datos que los modelos tradicionales.

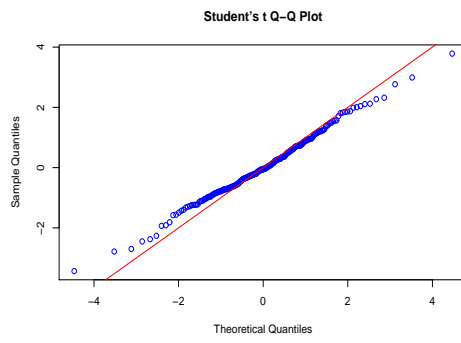
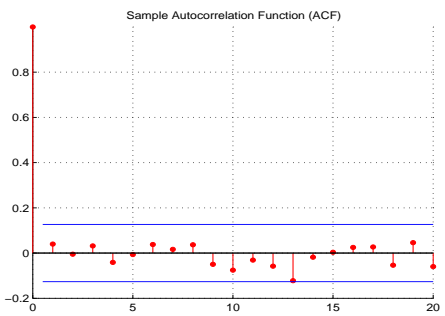
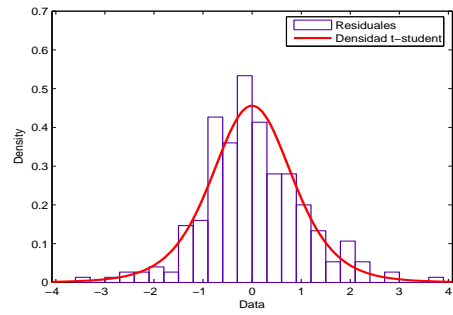
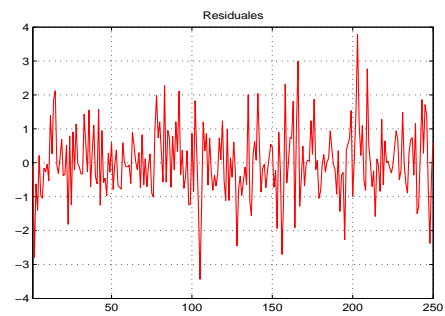


Figura 3.8: Residuales del modelo AR(1) - ARCH(1)

3.7.1. Predicción para observaciones futuras

A continuación se estudiará la efectividad del modelo para realizar predicciones tanto para los retornos como para los precios de la serie de datos. Lo que se hizo fue ajustar 20 modelos sustrayendo la última observación en cada modelo por separado, es decir, para el primero se consideraron $n - 1$ observaciones y se pronosticó la observación número n , para el segundo modelo se tomaron $n - 2$ datos y se pronosticó la observación $n - 1$ y así sucesivamente.

Para realizar las predicciones usamos la ventaja de que los retornos $\{Z_t\}_t$ sigue un proceso AR(1). Sabemos que si x_1, x_2, \dots, x_n es un conjunto de observaciones que obedecen un proceso AR(1), la predicción para una observación futura x_{n+1} puede obtenerse mediante la ecuación

$$\hat{x}_{n+1} = \hat{\phi}x_n, \quad (3.7)$$

El intervalo de confianza que presentaremos es análogo al intervalo de confianza para una predicción hecha mediante un modelo AR(1) pero adaptado a la distribución de los datos, es decir

$$\hat{x}_{inf} = x_{pred} - \hat{\beta} \times \xi_{t\nu}^{1-\alpha/2}, \quad \hat{x}_{sup} = x_{pred} + \hat{\beta} \times \xi_{t\nu}^{1-\alpha/2}, \quad (3.8)$$

donde \hat{x}_{inf} y \hat{x}_{sup} son los extremos inferior y superior del intervalo de confianza al $(1 - \alpha) \times 100$, $\hat{\beta}$ es el parámetro estimado para β , el parámetro de escala de la distribución *t-Student* de la densidad de los errores f_{Y_t} , y $\xi_{t\nu}^{1-\alpha/2}$ es el cuantil $(1 - \alpha) \times 100$ de una distribución $\mathcal{St}(0, 1, \nu)$. Los intervalos se construyeron con un nivel de significancia $\alpha = 0.05$.

Para realizar las predicciones de los precios podemos realizar estas a partir de las predicciones hechas para los retornos mediante el siguiente argumento. Sabemos que si definimos a los log-retornos como $r_t = \log(x_t/x_{t-1})$ para un conjunto de observaciones x_1, x_2, \dots , entonces la transformación inversa que tenemos que aplicar para regresar al dominio original de los datos es

$$x_t = x_{t-1}e^{r_t}.$$

Una vez calculados los intervalos de confianza para las predicciones de los retornos, podemos construir intervalos en el dominio de los precios con el argumento anterior de nueva cuenta.

A continuación se muestran las predicciones para los retornos y los precios junto con los valores verdaderos y los intervalo de confianza calculados al 95 %.

Predicción	Volar real	Extr. Inf.	Extr. Sup.
-0.107861	0.585247	-2.155472	1.939750
-0.407186	-0.615955	-2.464447	1.650076
-0.069076	-2.366615	-2.106771	1.968619
0.303591	-0.418911	-1.737108	2.344290
0.303952	1.777423	-1.733904	2.341809
0.104997	1.902481	-1.925434	2.135429
0.291737	0.683870	-1.747106	2.330580
0.013942	1.952818	-2.013806	2.041691
-0.210427	0.095465	-2.243336	1.822482
-0.169290	-1.422499	-2.201928	1.863348
0.179352	-1.216674	-1.848037	2.206741
-0.018449	1.216674	-2.045712	2.008815
0.140646	-0.125471	-1.890225	2.171516
0.114310	0.944889	-1.919755	2.148374
0.031833	0.794285	-2.004914	2.068580
-0.129142	0.223535	-2.169297	1.911013
-0.067159	-0.891158	-2.110278	1.975961
0.045275	-0.474159	-1.999863	2.090413
0.217922	0.315856	-1.835356	2.271201
-0.027437	1.530154	-2.077058	2.022185

Predicción	Volar real	Extr. Inf.	Extr. Sup.
32.335104	32.560000	31.679740	33.004027
32.437649	32.370000	31.777139	33.111888
33.326971	32.570000	32.654741	34.013039
33.591827	33.350000	32.913266	34.284377
33.000152	33.490000	32.334463	33.679547
32.313911	32.900000	31.664415	32.976729
32.153667	32.280000	31.504742	32.815959
31.444384	32.060000	30.813192	32.088505
31.343974	31.440000	30.713213	31.987690
31.806110	31.410000	31.166133	32.459228
32.307893	31.860000	31.659481	32.969584
31.854123	32.250000	31.214857	32.506480
31.944898	31.860000	31.302681	32.600290
31.636143	31.900000	30.999143	32.286231
31.359981	31.600000	30.727718	32.005254
31.239630	31.350000	30.608751	31.883513
31.538812	31.280000	30.900974	32.189815
31.724360	31.560000	31.082143	32.379847
31.678960	31.710000	31.035135	32.336141
31.121460	31.610000	30.490081	31.765914

Aparentemente las predicciones de los retornos y los precios de la acción no se asemejan mucho a las observaciones originales, sin embargo, se debe considerar que estamos trabajando con datos financieros en los cuales la volatilidad no permite dar una predicción puntual con una gran precisión. Considerando ahora las estimaciones por intervalos, podemos ver que solo en el tercer modelo el intervalo calculado no contuvo a la observación original, tanto en los retornos como en los precios; por tanto, 19 de los 20 intervalos calculados contuvieron a la observación original, lo cual es equivalente a decir que en el 95% de los casos el intervalo contuvo a la observación verdadera como se deseaba.

Notemos nuevamente que el parámetro β fue el resultado de usar variables latentes en el modelo y como vimos, juega un papel importante en la construcción del intervalo de confianza para las predicciones ya que, en otro caso, al usar un modelo tradicional este parámetro es considerado como 1 y esto

posiblemente hubiera ocasionado que los intervalos calculados no contuvieran a los verdaderos valores en más casos de los esperados.

3.7.2. Valuación de opciones y volatilidad estocástica

Una aplicación moderna muy interesante de la modelación de datos de algún activo financiero, a través de series de tiempo, es el poder utilizar el modelo ajustado para valorar opciones. Las opciones son instrumentos financieros que pertenece a una gama de productos llamados *derivados*, ya que su valor monetario se deriva o es función del precio de otro u otros activos financieros.

En particular, las opciones son un tipo de seguros financieros ya que brindan la opcionalidad (de ahí el nombre) de poder comprar o vender algún activo financiero a un precio pactado en una fecha futura. Si a la fecha de vencimiento de la opción, dadas las condiciones de mercado que haya en ese momento, si el tenedor de la opción se ve favorecido puede ejercer su derecho de comprar o vender el activo financiero a su contraparte; de ésta manera, éste obtendrá una ganancia monetaria. La analogía de los seguros con las opciones es que, para adquirir el derecho a comprar o vender un activo a un valor fijo, se tiene que pagar una prima en la cual debe estar reflejado el pago esperado que se tendría que hacer en caso de que el tenedor ejerza el derecho que adquirió, lo que en el caso de los seguros sería el pago por parte de la aseguradora en caso de presentarse una reclamación, derivada de la ocurrencia de un evento contingente.

De manera muy general podemos distinguir entre dos tipos de opciones; una opción que otorga al tenedor el derecho de compra (llamada *call*) y una opción que otorga al tenedor el derecho de venta (llamada *put*). En particular una opción de tipo *call*, ésta otorga el derecho mas no la obligación de comprar un activo financiero a un precio pactado, dentro de un tiempo determinado. Al final de este tiempo si conviene, se puede ejercer el derecho a comprar el activo al precio pactado o puede no ejercerse si no es conveniente, como ya se explicó. En las opciones llamadas *européas* este derecho sólo puede ejercerse hasta el vencimiento del instrumento y no antes, como es el caso de las opciones *americanas*.

Las opciones que poseen las características más sencillas en su contrato reciben el nombre genérico de opciones *plain vanilla* (véase Hull, 2005).

Uno de los objetivos principales dentro de la valuación de opciones es calcular la prima justa que debe cobrarse al vender una opción ya sea de tipo *call* o de tipo *put*. Gran parte de la investigación en economía y finanzas se ha centrado en desarrollar modelos que permitan una valuación correcta de las opciones entre los que destacan principalmente, para la valuación de opciones europeas, el modelo *Black-Scholes*, el método *Binomial* y el método de Monte Carlo. El primer y tercer modelo asumen que el precio del activo financiero, sobre el cual está suscrita la opción, evoluciona a tiempo continuo mientras que el segundo plantea que el precio evoluciona a tiempo discreto. Bajo ciertas condiciones estos 3 modelos de valuación convergen al mismo valor; para nosotros será de interés principal el método de Monte Carlo. Este método propone que el precio de un activo financiero puede modelarse a través de un proceso estocástico a tiempo continuo conocido como *Movimiento Browniano Geométrico*. La ecuación que describe la evolución en el tiempo de este proceso en general es

$$X(t + \delta t) = X(t)e^{(\mu - \frac{1}{2}\sigma^2)\delta t + \sigma\sqrt{\delta t}w}$$

donde $X(t)$ es el valor del proceso al tiempo t , μ es un parámetro de escala llamado deriva o tendencia, σ es otro parámetro llamado coeficiente de difusión y $w \sim N(0, 1)$. Los parámetros que inciden en el valor de una opción son el precio del activo financiero al tiempo t denotado $S(t)$, el valor pactado al cual puedo comprar o vender, conocido como *strike* o *precio de ejercicio* y denotado K , una tasa de interés compuesta continuamente y libre de riesgo r , el tiempo de vida que dura la opción denotado $T - t$ y el factor más importante en nuestra aplicación es que la volatilidad del activo financiero denotada por σ . Si la opción se inicia al tiempo $t = 0$, el precio actual del activo es $S(0)$ y la duración de la opción es T ; de esta manera si la adaptamos a la ecuación del movimiento Browniano Geométrico obtendremos un modelo que describa el precio del activo financiero en tiempo continuo en función de los parámetros ya mencionados de la siguiente manera

$$S(T) = S(0)e^{(r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}w}, \quad (3.9)$$

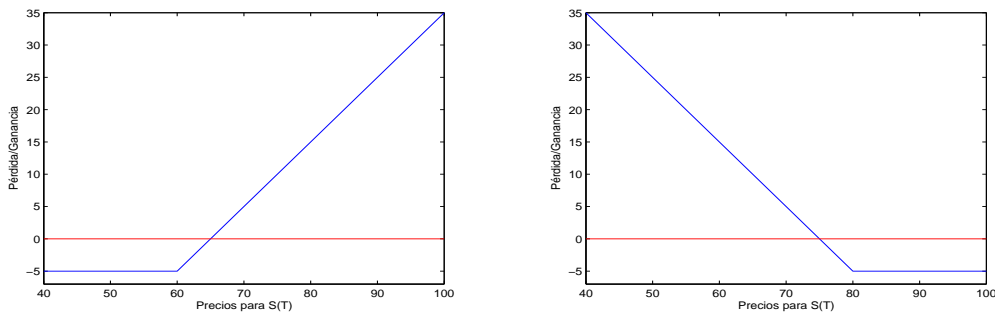


Figura 3.9: Diagramas de pérdidas y ganancias para una opción *call* (izquierda) y *put* (derecha) con *strikes* de 60 y 80 respectivamente

donde para este caso $\mu = r$ y $\delta t = T$. El método Monte Carlo usa esta ecuación para obtener suficientes muestras de $S(T)$ y a su vez suficientes muestras del pago que se hará (llamado *payoff*) en caso de que el tenedor ejerza su derecho. El valor de la opción será entonces el promedio de los *payoffs* obtenidos mediante la simulación descontados con la tasa de interés r . Para el caso de una opción de *call* el *payoff* que se recibirá al tiempo de vencimiento T será

$$\text{máx}(S(T) - K, 0)$$

y análogamente para una opción de tipo *put*

$$\text{máx}(K - S(T), 0)$$

donde $S(T)$ es el precio del subyacente al tiempo T al cual también denotaremos como S_T y K el precio de ejercicio pactado o simplemente el *strike* (véase Hull, 2005).

Para una mayor comprensión de estos conceptos en la figura 3.9 se muestra un diagrama de utilidades y pérdidas para el caso de ambas opciones *call* y *put*.

En este ejemplo vemos que el costo de ambas primas es 5 y dependiendo de los valores que tome S_T se obtendrá una pérdida o una ganancia. Es por

ello que un método intuitivo para obtener el valor que debe cobrarse por la opción es generar muchos valores para S_T como hace el método Monte Carlo. Regresando al análisis de este método vemos que el valor de la prima c para una opción *call* mediante el método Monte Carlo puede obtenerse mediante la expresión

$$c = \frac{1}{n} \sum_{i=1}^n \text{máx}(S_{T,i} - K, 0),$$

y cada muestra $S_{T,i}$ es obtenida a través de simulación del modelo (3.9).

Existen otro tipo de opciones cuyo modelo de valuación es más complejo debido a las características que posee. Éstas reciben el nombre genérico de *opciones exóticas path-depending* y el valor de éstas depende no sólo del valor S_T sino de toda la trayectoria que describió el precio del activo durante el período de vida de la opción. Una forma de generar posibles trayectorias para el precio de algún activo es nuevamente a través del movimiento browniano geométrico pero haciendo una partición en el tiempo $\delta t = T/m$, donde T es el período de vida de la opción y m la partición que se hace en el tiempo; conforme $m \rightarrow \infty$, la simulación computacional se asemeja más a un proceso que evoluciona a tiempo continuo como se sucede en la realidad con los precios de los activos financieros. De esta manera la ecuación (3.9) puede modificarse como

$$S(t) = S(t-1)e^{(r-\frac{1}{2}\sigma^2)\delta t + \sigma\sqrt{\delta t}w}, \quad (3.10)$$

para $t = 1, \dots, m$. En la figura 3.10 podemos ver un gráfico donde se muestran 100 trayectorias simuladas a partir de la ecuación (3.10). Ésta es la base para idear un modelo que sirva para valorar opciones a partir de la trayectoria que describe desde un tiempo inicial t hasta un vencimiento T . Antes es conveniente citar el conocido modelo de *Black-Scholes*, el cual es un método analítico para valorar opciones del tipo europeo *plain vanilla*. Esta ecuación asume que el proceso estocástico que describe la evolución del precio del activo financiero en el tiempo es el movimiento browniano geométrico. En el año 1997 Fisher Black y Myron Scholes recibieron el premio Nobel de economía por sus investigaciones en las que desarrollaron el siguiente modelo para valorar una opción sobre un activo financiero que no paga dividendos:

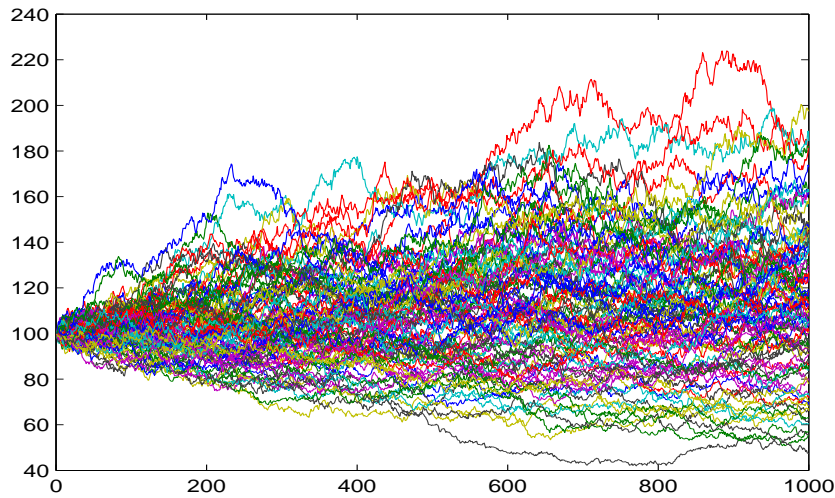


Figura 3.10: Simulación de 100 trayectorias de precios con $m=1000$.

$$c = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2),$$

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 = d_1 - \sigma\sqrt{T-t}$$

y $N(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

En particular ésta es la fórmula para valorar un *call*; para un put la fórmula se modifica. Este modelo actualmente es usado como un método analítico que calcula el valor correcto que debe pagarse por una opción aunque hay dos supuestos muy fuertes que asume y los cuales no se apegan a la realidad

- (1) Como consecuencia de modelar el precio del activo financiero de la opción a través del movimiento browniano geométrico, estipula que los retornos de $S(t)$ tiene distribución normal, supuesto que no se cumple

en la práctica ya que por lo regular las distribuciones que mejor modelan los retornos comúnmente poseen colas más pesadas que las de la distribución normal.

- (2) Considera que la volatilidad σ es constante a través del tiempo lo cual tampoco llega a cumplirse, tanto que debido a la heterocedasticidad de los retornos de activos financieros surgió el interés por desarrollar los modelos ARCH y GARCH.

Tomando todo esto en cuenta, se desarrolló un método para valuar una opción tipo *call* sobre la acción *ATT Inc.*, cuyos retornos se modelaron en la sección anterior, simulando posibles trayectorias futuras para un plazo de tiempo determinando a partir de la simulación de sus retornos, los cuales suponemos siguen un proceso AR(1)-ARCH(1). En términos matemáticos, sean y_t y S_t la volatilidad y el precio de la acción al tiempo t respectivamente. Entonces dados los retornos x_n, x_{n-1} y el precio S_n se tiene que

$$y_n = x_n - \phi x_{n-1}.$$

Luego para $t = n + 1, \dots, \tau$

$$y_t | w_{t-1} \sim N(0, w_{t-1})$$

$$x_t = \hat{\phi} x_{t-1} + y_t$$

$$w_t | y_t \sim \mathcal{I}g((\hat{\nu} + 1)/2, (x_t - \hat{\phi} x_{t-1})^2/2 + (\hat{\nu} \hat{\beta}^2/2))$$

$$S_t = S_{t-1} e^{x_t}$$

En la figura 3.11 se muestra un ejemplo de trayectorias simuladas con nuestro modelo.

A continuación se muestra un ejemplo de lo descrito anteriormente para los siguientes datos: S_t = último precio observado en los datos, $K = 30$, $r = .1$ y $T = 40$ días. El número de simulaciones que se realizaron para la valuación fueron 10,000

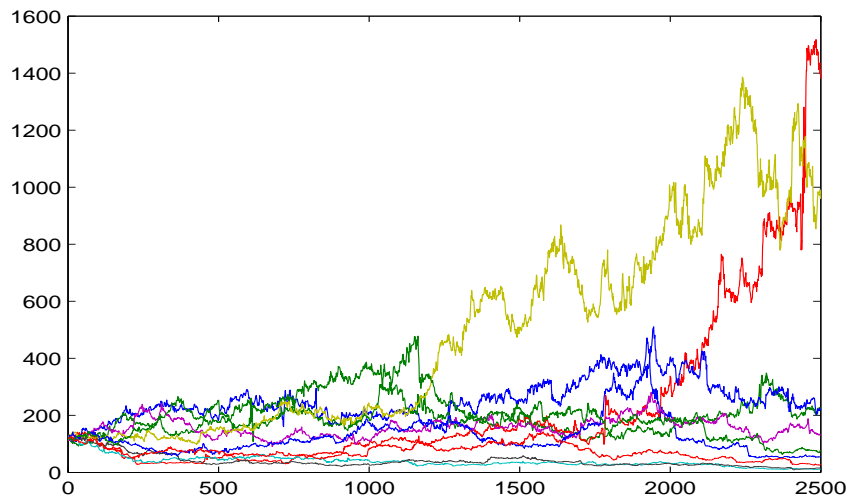


Figura 3.11: Simulación de 10 trayectorias de precios con retornos AR-ARCH.

```
>> [sim_val,bls_price]=pricing_call(St,30,r,40,...
vol_bls*sqrt(360),beta,nu,phi,sim,xn1,xn2)
```

```
sim_val =
```

```
2.9771
```

```
bls_price =
```

```
2.9615
```

```
>> binom(St,K,r,d/360,vol_bls*sqrt(360),3)
```

```
Call
```

```
2.9798
```

Lo que se muestra arriba es el resultado de la implementación en Matlab comparado con el precio calculado con la fórmula de Black-Scholes y el precio

usando el método binomial con 3 períodos. Como vemos, los 3 precios son similares sobre todo el de la simulación de nuestro modelo y el que se obtiene con el método binomial. Podemos pensar que se tiene una mayor diferencia con la fórmula de Black-Scholes debido a que hemos confirmado que los datos no presentan una volatilidad constante en el tiempo además de que los retornos no tiene distribución Normal como son las hipótesis de dicha fórmula.

Podemos considerar que el precio obtenido mediante la simulación de los retornos que se modelaron refleja a mayor detalle la estructura estocástica, tanto de los retornos como de los precios comparado con las otras dos alternativas. Sin embargo, no se puede concluir que esta valuación sea mejor que las otras dos, sólo es un ejemplo ilustrativo de las aplicaciones que pudieran realizarse al modelar retornos como series de tiempo con la estructura que se plantea. Para un mayor detalle de estos conceptos existe una gran variedad de libros con tópicos de derivados financieros siendo dentro de ellos el más representativo el libro de Hull (2005).

3.7.3. Estimación del VaR

En un contexto financiero denotemos al proceso X_t como los retornos de algún activo financiero. Asumamos que X_t obedece un modelo GARCH (ó ARMA-GARCH) y a $f(\cdot)$ como la función de densidad del error ε_t . Con frecuencia en aplicaciones financieras es de interés conocer algún cuantil extremo de esta distribución, el cual es llamado *valor en riesgo* (VaR). Para $\alpha \in (0, 1)$, el $100\alpha\%$ cuantil (condicional) está definido como

$$x_\alpha = \inf\{x : \mathbb{P}(X_t \leq x | X_{t-k}, k \geq 1) \geq \alpha\}.$$

Un x_α cuantil extremo alto con α muy cercano a 1 (o un cuantil extremo bajo con α muy cercano a 0) representa una pérdida potencial con probabilidad α (ó $1 - \alpha$). El VaR es la medida más utilizada en la administración de riesgos financieros. Se elaboró un ejemplo sobre esta aplicación en Matlab comparada con el método conocido como *paramétrico* suponiendo primeramente una distribución Normal, como se hace cotidianamente en este método (véase Hull, 2005), y también considerando una distribución *t-Student*.

A continuación se presentan los comandos en Matlab para llevar a cabo el cálculo mencionado.

```
>> x0=icdf('tlocationscale',0.05,0,beta,nu)
```

```
x0 =
```

```
-1.6550
```

```
>> param=std(r_att,1)*norminv(.05,0,1)
```

```
param =
```

```
-1.6956
```

```
>> param2=std(r_att,1)*icdf('tlocationscale',0.05,0,1,nu)
```

```
param2 =
```

```
-1.8994
```

Vemos que nuestro cálculo no se aleja mucho del obtenido mediante el método paramétrico considerando una distribución Normal incluso siendo más parecido éste que el que considera una distribución *t-Student*. Aun así, sabemos que los retornos son más cercanos a provenir de una distribución *t-Student* reescalada (como la que planteamos) que de una Normal o una *t-Student* estandarizada, por lo cual nuestra estimación para el *VaR* es un buen candidato para acotar o predecir futuras pérdidas financieras.

Capítulo 4

Conclusiones

A lo largo de este trabajo hemos aprendido, si no un algoritmo genérico, un ejemplo en donde se ilustra cómo la introducción de variables latentes permite definir la relación de dependencia estocástica en un modelo estadístico, en este caso una serie de tiempo del tipo AR(1)-ARCH(1), siendo ésta una de muchas aplicaciones que puede haber en la utilización de estas técnicas. Vimos cómo la estadística Bayesiana ha motivado el desarrollo de algoritmos intensivos de simulación estocástica de los cuales surge la idea de la incorporación de variables latentes como observaciones auxiliares en cualquier modelo probabilístico.

Además se ilustra de manera particular el uso de una distribución marginal específica para un conjunto de observaciones en estudio, rompiendo con el esquema tradicional del uso excesivo de la distribución normal para gobernar la estructura probabilística de los datos de interés. Si bien en esta tesis se dio tratamiento a una colección de observaciones pertenecientes a un activo financiero, puede aplicarse esta misma idea para cualquier análisis de datos que permitan modelarse a través de series de tiempo, ya sea considerando una varianza condicional homocedástica o heterocedástica como fue en nuestro caso, e incorporando una estructura de dependencia temporal entre las observaciones, tanto en la media como en la varianza condicional.

Cabe mencionar que, en el caso expuesto en el trabajo, sólo se introdujo una variable latente que permitía definir de manera particular la relación de dependencia entre las observaciones; sin embargo es factible introducir 2 o más variables latentes en el modelo según convenga y permita definir completamente la relación de dependencia estocástica de las observaciones fijando

una distribución marginal apropiada, además de preservar la estacionariedad de las observaciones.

Como consecuencia de usar la idea de crear dependencia vía variables latentes, nuestro modelo AR(1)-ARCH(1) tiene errores con distribución marginal $\mathcal{St}(0, \beta^2, \nu)$. Los modelos ARMA-GARCH implementados en varias paqueterías como Matlab, consideran una distribución $\mathcal{St}(0, 1, \nu)$, la cual resulta menos flexible. Consideremos que esta es una ventaja de usar nuestro esquema.

Ahora bien, pensemos que definiendo las distribuciones de transición apropiadas entre las observaciones y las variables latentes podemos, mediante el teorema de Bayes, especificar cualquier distribución de probabilidad, continua o discreta, como distribución marginal de nuestras observaciones. A su vez, puede considerarse más de un proceso latente inherente a la serie observable y de esta manera, construir modelos de series de tiempo cada vez más complejos como pudo ser el caso de considerar un proceso GARCH para los errores en el modelo propuesto en vez de un proceso ARCH, en cuyo caso se hubiera tenido que hacer uso de una segunda variable latente como se hace en el artículo de Pitt y Walker (2005); lo anterior también requeriría del uso de métodos de estimación alternativos al algoritmo E-M como son los métodos MCMC.

Apéndice A

Algoritmo Gibbs sampler

Las técnicas conocidas en el nombre genérico de Monte Carlo vía Cadenas de Markov (MCMC) permiten generar observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos directos. Estos métodos son de gran utilidad en la Inferencia Bayesiana ya que podemos conocer características de la distribución *a posteriori* a pesar de que ésta sea difícil de obtener, después de la aplicación del teorema de Bayes.

La idea básica es construir una cadena de Markov que sea fácil de simular y cuya distribución estacionaria corresponda a la distribución *a posteriori* que nos interesa. Otras características que se piden a la cadena es que sea homogénea, irreducible y aperiódica (véase Gamerman, 1997, Robert y Casella 1999) .

Una cuestión que debe abordarse es qué tan sensible es la cadena al estado inicial X_0 . Dadas las condiciones mencionadas en la Sección 1.1.1, la cadena “se olvidará” de su estado inicial y convergerá una distribución estacionaria $\pi(\mathbf{x})$.

Supongamos que hemos producido n muestras de la cadena siendo n un número suficientemente grande como para que haya alcanzado la convergencia enunciada para un tiempo $m + 1$ con $m + 1 < n$. Asumimos que las muestras \mathbf{X}_t , $t = m + 1, \dots, n$ se distribuyen de acuerdo a la distribución estacionaria $\pi(\mathbf{x})$. Se desechan las primeras m iteraciones y se usan las $n - m$ muestras restantes entonces estimamos la esperanza de la siguiente manera

$$E[f(\mathbf{X})] = \frac{1}{n-m} \sum_{t=m+1}^n f(\mathbf{X}_t).$$

Cuando se desecha un número m de muestras, se dice que se “quemar” las primeras m iteraciones de la cadena y a m se le conoce como “período de calentamiento”. El valor de m y n es un tema actual en investigación de métodos MCMC. El texto de Geyer (1992) sugiere que las observaciones que se deben quemar son entre el 1% y el 2% de n , aunque cada caso particular requiere establecer un número m adecuado.

El problema a tratar a continuación, se refiere a cómo construir una cadena de Markov $\{\theta_t\}_t$ que cumpla con las hipótesis necesarias para la existencia de su distribución estacionaria $\pi(\cdot)$, pero además, que sea tal que $\pi(\theta)$ coincida con la distribución posterior $\mathbb{P}(\theta|\mathbf{x})$

El algoritmo conocido como Gibbs sampler forma parte de los métodos de MCMC. Para el caso de una distribución $f(\mathbf{x})$ definida en \mathbb{R}^k , este algoritmo permite simular muestras de una cadena de Markov $x^{(1)}, x^{(2)}, \dots$ con distribución estacionaria o de equilibrio $f(\mathbf{x})$. Para lo anterior se requieren las densidades condicionales completas

$$\begin{aligned} & f(x_1|x_2, \dots, x_n) \\ & f(x_2|x_1, x_3, \dots, x_n) \\ & \vdots \\ & f(x_i|x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \\ & \vdots \\ & f(x_n|x_1, x_2, \dots, x_{n-1}). \end{aligned} \tag{A.1}$$

En general, estas densidades no siempre son fáciles de obtener o bien, puede no ser fácil producir muestras de ellas, lo cual es crucial para el algoritmo Gibbs Sampler. Por las razones anteriores, en ocasiones es preferible utilizar un algoritmo más general, conocido como *Metropolis-Hastings*.

Una aplicación de este método se da cuando tenemos una densidad conjunta dada por $f(x, y_1, \dots, y_n)$ y nos gustaría conocer ciertas características acerca de la densidad marginal dada por

$$f(x) = \int \cdots \int f(x, y_1, \dots, y_n) dy_1 \dots dy_n.$$

En la ecuación de arriba digamos que queremos obtener la distribución marginal, debemos integrar entonces sobre todas las otras variables. En muchas aplicaciones, esta integración puede ser muy difícil y algunas veces imposible de realizar analíticamente. El Gibbs sampler nos provee con una manera de obtener $f(x)$ a través de simulación.

Una forma de hacer ésto será considerar que

$$f(x, y_1, \dots, y_k) = f(x|y_1, \dots, y_k)f(y_1, \dots, y_k)$$

luego entonces

$$f(x) = \int \dots \int f(x|y_1, \dots, y_k)f(y_1, \dots, y_k)dy_1 \dots dy_k.$$

En tal situación, podríamos usar el Gibbs Sampler para obtener muestras $(y_1^{(i)}, \dots, y_k^{(i)})$ $i = 1, 2, \dots, n$ de $f(y_1, \dots, y_k)$ y luego usar el método de Monte Carlo descrito en la sección anterior para aproximar $f(x)$ por

$$\frac{1}{n} \sum_{i=1}^n f(x|y_1^{(i)}, \dots, y_k^{(i)}).$$

Alternativamente, en caso de poder usar los métodos MCMC para simular directamente muestras $(x^{(i)}, y_1^{(i)}, \dots, y_k^{(i)})$, $i = 1, 2, \dots, n$ de $f(x, y_1, \dots, y_k)$, entonces la muestra $\{x^{(i)}\}_{i=1}^n$ correspondería a la densidad marginal $f(x)$.

Para ilustrar el Gibbs sampler, comenzaremos por ver un caso simple donde la distribución conjunta es $f(x_1, x_2)$. Usando la notación de la sección previa, \mathbf{X}_t es un vector bidimensional con elementos dados por

$$\mathbf{X}_t = (x_{t,1}, x_{t,2}).$$

Dados los valores iniciales $\mathbf{X}_0 = (x_{0,1}, x_{0,2})$, generamos una muestra de $f(x_1, x_2)$ al muestrear x_1 de la distribución condicional dada por $f(x_1|x_2 = x_{0,2})$ y x_2 de la distribución condicional $f(x_2|x_1)$. En cada iteración, las entradas de \mathbf{X}_t son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Se ilustra esto en el procedimiento dado abajo.

ALGORITMO GIBBS SAMPLER (CASO BIVARIADO)

Para la t -ésima iteración, dada la muestra $\mathbf{X}_t = (x_{t,1}, x_{t,2})$

1. Generamos un valor $x_{t+1,1}$ de

$$f(X_1|X_2 = x_{t,2}). \quad (\text{A.2})$$

2. Generamos un valor $x_{t+1,2}$ de

$$f(X_2|X_1 = x_{t+1,1}). \quad (\text{A.3})$$

3. La $(t + 1)$ -ésima muestra de la cadena es $\mathbf{X}_{t+1} = (x_{t+1,1}, x_{t+1,2})$.

Notemos que las distribuciones (A.1) o bien (A.2) y (A.3) están condicionadas en el actual o más reciente valor de las componentes de \mathbf{X}_t .

Al repetir los pasos 1,2 y 3 un número n de veces grande, los valores $(x_{t+m,1}, x_{t+m,2}), (x_{t+m+1,1}, x_{t+m+1,2}), \dots$, para $m \in \mathbb{Z}, 0 \leq m \leq n$ serán una muestra aproximadamente de $f(x_1, x_2)$. El valor del período de calentamiento m se selecciona de forma que la cadena de Markov $\mathbf{X}_{t+1}, \mathbf{X}_{t+2}, \dots$ haya alcanzado su distribución estacionaria, a saber $f(x_1, x_2)$.

Ejemplo 9 *Sólo con el fin de ilustrar lo anterior, usaremos el Gibbs sampler para generar muestras de una normal bivariada con los siguientes parámetros*

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

Del texto de Gelman (1995) sabemos que $f(x_1|x_2)$ es normal univariada con media $\mu_1 + \rho(x_2 - \mu_2)$ y desviación estándar $1 - \rho^2$. Similarmente, $f(x_2|x_1)$ es normal univariada con media $\mu_2 + \rho(x_1 - \mu_1)$ y desviación estándar $1 - \rho^2$. Con esta información, podemos implementar el Gibbs sampler para generar las variables aleatorias.

En la figura A.1 se muestra un “scatter plot” (o diagrama de dispersión) de la simulación de las variables normales bivariadas producidas al implementar el Gibbs sampler. Para más detalles sobre los métodos MCMC se

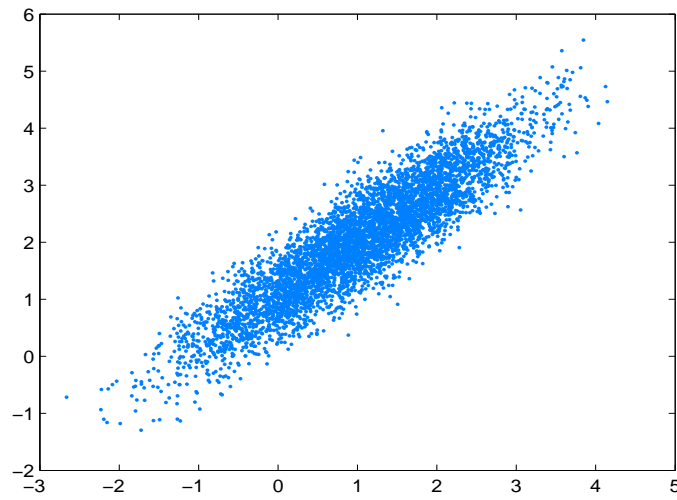


Figura A.1: Gráfico de dispersión de las variables generadas por el Gibbs sampler

pueden consultar Gutiérrez-Peña (1997) o Gamerman (1997). Lo que veremos a continuación es un método relacionado con el Gibbs sampler conocido como variables latentes, el cual es uno de los temas de mayor importancia de esta tesis.

Apéndice B

Teoremas y cálculos

B.1. Teorema de convergencia monótona (TCM)

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Sean X_1, X_2, \dots una sucesión de v.a. en $(\Omega, \mathcal{F}, \mathbb{P})$, no negativas y tales que $X_n \leq X_{n+1}$. Sea

$$X(\omega) \equiv \lim_{n \rightarrow \infty} X_n(\omega), \quad \text{para cada } \omega \in \Omega.$$

Entonces

$$\mathbb{E}(X_n) \rightarrow \mathbb{E}(X) \quad \text{si } n \rightarrow \infty.$$

En resumen: $0 \leq X_n \uparrow X$ implica que $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ cuando $n \rightarrow \infty$.

B.2. Teorema de convergencia dominada (TCD)

Corolario: Si X_1, X_2, \dots son v.a. no negativas

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n).$$

Teorema 5 (*Teorema de Convergencia Dominada*)

Sean $X_1, X_2 \dots X$ y Y variables aleatorias en $(\Omega, \mathcal{F}, \mathbb{P})$. Supóngase que $|X_n| \leq Y$, para toda $n \geq 1$ y que $\mathbb{E}(Y) < +\infty$. Si además

$$X = \lim_{n \rightarrow \infty} X_n$$

con probabilidad 1 (ó \mathbb{P} c.s.), entonces $\mathbb{E}(|X|) < +\infty$ y

$$\mathbb{E}(X_n) \rightarrow \mathbb{E}(X) \quad \text{cuando } n \rightarrow \infty.$$

B.3. Estimación de parámetros para el proceso ARCH(1)

La verosimilitud de α_0 y α_1 es

$$\begin{aligned} L(\alpha_0, \alpha_1 | y_1) &= \prod_{t=2}^n f_{\alpha_0, \alpha_1}(y_t | y_{t-1}) \\ &= \prod_{t=2}^n \frac{1}{\sqrt{2\pi(\alpha_0 + \alpha_1 y_{t-1}^2)}} e^{-\frac{1}{2(\alpha_0 + \alpha_1 y_{t-1}^2)} y_t^2} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n-1}{2}} \prod_{t=2}^n \frac{1}{\sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}} e^{-\frac{1}{2} \sum_{t=2}^n \frac{y_t^2}{2(\alpha_0 + \alpha_1 y_{t-1}^2)}} \end{aligned}$$

$$l(\alpha_0, \alpha_1) \propto -\log L(\alpha_0, \alpha_1 | y_1)$$

$$l(\alpha_0, \alpha_1) = \left(\frac{1}{2}\right) \sum_{t=2}^n \log(\alpha_0 + \alpha_1 y_{t-1}^2) + \left(\frac{1}{2}\right) \sum_{t=2}^n \left\{ \frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2} \right\}$$

$$\begin{aligned} \frac{\partial l}{\partial \alpha_0} &= \left(\frac{1}{2}\right) \sum_{t=2}^n \frac{1}{\alpha_0 + \alpha_1 y_{t-1}^2} - \left(\frac{1}{2}\right) \sum_{t=2}^n y_t^2 \frac{1}{(\alpha_0 + \alpha_1 y_{t-1}^2)^2} \\ &= \sum_{t=2}^n \frac{\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2}{2(\alpha_0 + \alpha_1 y_{t-1}^2)^2} \end{aligned}$$

$$\begin{aligned}
\frac{\partial l}{\partial \alpha_1} &= \left(\frac{1}{2}\right) \sum_{t=2}^n \frac{y_{t-1}^2}{\alpha_0 + \alpha_1 y_{t-1}^2} - \left(\frac{1}{2}\right) \sum_{t=2}^n y_t^2 \frac{y_{t-1}^2}{(\alpha_0 + \alpha_1 y_{t-1}^2)^2} \\
&= \sum_{t=2}^n y_{t-1}^2 \left[\frac{\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2}{2(\alpha_0 + \alpha_1 + y_{t-1}^2)^2} \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \alpha_1 \partial \alpha_0} &= \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2)^2 (y_{t-1}^2) - (\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2) 2(\alpha_0 + \alpha_1 y_{t-1}^2) (y_{t-1}^2)}{(\alpha_0 + \alpha_1 y_{t-1}^2)^4} \\
&= \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2) (y_{t-1}^2) [\alpha_0 + \alpha_1 y_{t-1}^2 - 2\alpha_0 - 2\alpha_1 y_{t-1}^2 + 2y_t^2]}{(\alpha_0 + \alpha_1 y_{t-1}^2)^4} \\
&= - \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(y_{t-1}^2) [\alpha_0 + \alpha_1 y_{t-1}^2 - 2y_t^2]}{(\alpha_0 + \alpha_1 y_{t-1}^2)^3}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \alpha_0^2} &= \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2)^2 - (\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2) 2(\alpha_0 + \alpha_1 y_{t-1}^2)}{(\alpha_0 + \alpha_1 y_{t-1}^2)^4} \\
&= \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2) [\alpha_0 + \alpha_1 y_{t-1}^2 - 2\alpha_0 + 2\alpha_1 y_{t-1}^2 + 2y_t^2]}{(\alpha_0 + \alpha_1 y_{t-1}^2)^4} \\
&= - \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{\alpha_0 + \alpha_1 y_{t-1}^2 - 2y_t^2}{(\alpha_0 + \alpha_1 y_{t-1}^2)^3}
\end{aligned}$$

$$\frac{\partial^2 l}{\partial \alpha_0 \partial \alpha_1} = - \sum_{t=2}^n \left(\frac{1}{2}\right) \frac{(y_{t-1}^2) [\alpha_0 + \alpha_1 y_{t-1}^2 - 2y_t^2]}{(\alpha_0 + \alpha_1 y_{t-1}^2)^3}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \alpha_1^2} &= \sum_{t=2}^n \left(\frac{y_{t-1}^2}{2}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2)^2 (y_{t-1}^2) - (\alpha_0 + \alpha_1 y_{t-1}^2 - y_t^2) 2(\alpha_0 + \alpha_1 y_{t-1}^2) y_{t-1}^2}{(\alpha_0 + \alpha_1 y_{t-1}^2)^4} \\
&= - \sum_{t=2}^n \left(\frac{y_{t-1}^2}{4}\right) \frac{(\alpha_0 + \alpha_1 y_{t-1}^2 - 2y_t^2)}{(\alpha_0 + \alpha_1 y_{t-1}^2)^3}
\end{aligned}$$

B.4. Cálculos de la transición del proceso del Capítulo 4

Como vimos en el capítulo 4 $f_{Y_t, W_t}(y, w) = f_{Y_t}(y)f_{W_t|Y_t}(w|y)$, luego

$$\begin{aligned}
 f_{W_t}(w) &= \int f_{Y_t, W_t}(y, w) dy = \int f_{Y_t}(y) \times f_{W_t|Y_t}(w|y) dy \\
 &= \int_{-\infty}^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi} \right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{y^2}{\nu\beta^2} \right]^{\frac{\nu+1}{2}}} \times \frac{\left(\frac{\nu\beta^2}{2} + \frac{y^2}{2} \right)^{\frac{\nu+1}{2}}}{\Gamma(\frac{\nu+1}{2})} w^{-(\frac{\nu+1}{2}+1)} \\
 &\times e^{-\frac{\left(\frac{\nu\beta^2}{2} + \frac{y^2}{2} \right)}{w}} dy \\
 &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi} \right)^{\frac{1}{2}} \frac{1}{\Gamma(\frac{\nu+1}{2})} w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\nu\beta^2}{2w}} \int_{-\infty}^{\infty} \frac{1}{\left[1 + \frac{Y_t^2}{\nu\beta^2} \right]^{\frac{\nu+1}{2}}} \left(\frac{\nu\beta^2}{2} + \frac{y^2}{2} \right)^{\frac{\nu+1}{2}} \\
 &\times e^{-\frac{y^2}{2w}} dy \\
 &= \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi} \right)^{\frac{1}{2}} w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\nu\beta^2}{2w}} \int_{-\infty}^{\infty} \frac{1}{\left[1 + \frac{y^2}{\nu\beta^2} \right]^{\frac{\nu+1}{2}}} \left(\frac{\nu\beta^2}{2} \right)^{\frac{\nu+1}{2}} \left(1 + \frac{y^2}{\nu\beta^2} \right)^{\frac{\nu+1}{2}} \\
 &\times e^{-\frac{y^2}{2w}} dy \\
 &= \frac{\left(\frac{\nu\beta^2}{2} \right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{2w}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi w}} e^{-\frac{y^2}{2w}} dy \\
 &= \frac{\left(\frac{\nu\beta^2}{2} \right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{2w}}.
 \end{aligned}$$

Entonces concluimos que $f_{W_t}(w) = \mathcal{I}g(\frac{\nu}{2}, \frac{\nu\beta^2}{2})$. Ahora, si aplicamos el teorema de Bayes, tenemos que

$$\begin{aligned}
f_{Y_t|W_t}(y_t|w) &= \frac{f_{Y_t, W_t}(y_t, w)}{f_{W_t}(w)} \\
&= \frac{f_{Y_t}(y_t) \times f_{W_t|Y_t}(w|y_t)}{f_{W_t}(w)} \\
&= \frac{\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi}\right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{(y_t)^2}{\nu\beta^2}\right]^{\frac{\nu+1}{2}}} \times \frac{\left(\frac{\nu\beta^2 + (y_t)^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma(\frac{\nu+1}{2})} w^{-(\frac{\nu+1}{2}+1)}}{\frac{(\frac{\nu\beta^2}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{w}}} \\
&\times e^{-\frac{\left(\frac{\nu\beta^2 + (y_t)^2}{2}\right)}{w}} \\
&= \frac{\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu\beta^2\pi}\right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{(y_t)^2}{\nu\beta^2}\right]^{\frac{\nu+1}{2}}} \times \frac{\frac{\nu\beta^2}{2}^{\frac{\nu+1}{2}} \left(1 + \frac{(y_t)^2}{\nu\beta^2}\right)^{\frac{\nu+1}{2}}}{\Gamma(\frac{\nu+1}{2})} w^{-(\frac{\nu+1}{2}+1)}}{\frac{(\frac{\nu\beta^2}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{w}}} \\
&\times e^{-\frac{\left(\frac{\nu\beta^2 + (y_t)^2}{2}\right)}{w}} \\
&= \frac{\left(\frac{1}{\nu\beta^2\pi}\right)^{\frac{1}{2}} \times \frac{\nu\beta^2}{2}^{\frac{\nu+1}{2}} w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\left(\frac{\nu\beta^2 + (y_t)^2}{2}\right)}{w}}}{\left(\frac{\nu\beta^2}{2}\right)^{\frac{\nu}{2}} w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{w}}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{(\nu\beta)^{\frac{\nu+1}{2}} \left(\frac{1}{\pi^{\frac{1}{2}}}\right) \left(\frac{1}{2^{\frac{\nu+1}{2}}}\right) w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)}{w}}}{(\nu\beta)^{\frac{\nu+1}{2}} \left(\frac{1}{2^{\frac{\nu}{2}}}\right) w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{2w}}} \\
&= \frac{\left(\frac{1}{\pi^{\frac{1}{2}}}\right) \left(\frac{1}{2^{\frac{1}{2}}}\right) w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)}{w}}}{w^{-(\frac{\nu}{2}+1)} e^{-\frac{\nu\beta^2}{2w}}} \\
&= \left(\frac{1}{\pi^{\frac{1}{2}}}\right) \left(\frac{1}{2^{\frac{1}{2}}}\right) w^{-(\frac{1}{2}+1)} e^{-\frac{\left(\frac{(y_t)^2}{2}\right)}{w}} \\
&= \frac{1}{\sqrt{2\pi w}} e^{-\frac{1}{2w}(y_t)^2}
\end{aligned}$$

Por lo tanto $f_{Y_t|W_t}(y_t|w) = N(0, w)$. Ahora veamos la forma que tiene la transición de $\{Y_t\}_t$. Usando los resultados anteriores tenemos que

$$\begin{aligned}
f_{W_t|Y_t}(w|y_t) &= \frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} w^{-(\frac{\nu+1}{2}+1)} e^{-\frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)}{w}} \\
f_{Y_{t+1}|W_t}(y_{t+1}|w) &= \frac{1}{\sqrt{2\pi w}} e^{-\frac{1}{2w}(y_{t+1})^2}.
\end{aligned}$$

Entonces

$$f_{Y_{t+1}|Y_t}(y_{t+1}|y_t) = \int_0^\infty f_{Y_{t+1}|W_t}(y_{t+1}|w) f_{W_t|Y_t}(w|y_t) dw$$

$$\begin{aligned}
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} w^{-(\frac{\nu+1}{2} + \frac{3}{2})} e^{-\frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2} + \frac{(y_{t+1})^2}{2}\right)}{w}} dw \\
&= \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\left(\frac{\nu\beta^2}{2} + \frac{(y_t)^2}{2} + \frac{(y_{t+1})^2}{2}\right)^{\frac{\nu+2}{2}}} \\
&= \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} \left(\frac{1}{\pi(\nu\beta^2 + (y_t)^2)}\right)^{\frac{1}{2}} \frac{1}{\left[1 + \frac{(y_{t+1})^2}{\nu\beta^2 + y_t^2}\right]^{\frac{\nu+2}{2}}}
\end{aligned}$$

con lo cual hemos verificado que $f_{Y_{t+1}|Y_t}(y_{t+1}|y_t)$ es una densidad

$$St(y_{t+1}|0, \lambda = \frac{\nu+1}{\nu\beta^2 + (y_t)^2}, \alpha = \nu+1).$$

B.5. Cálculos del Paso-E del Capítulo 4

Un cálculo pendiente en el *Paso-E* era resolver

$$\int_0^\infty \log(w) w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw \quad (\text{B.1})$$

$$\text{con } a^{(k)} = \frac{\nu^{(k)}+2}{2}, \quad b^{(k)} = \frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{2}.$$

Haciendo el cambio de variable $v = \frac{1}{w} \rightarrow w = \frac{1}{v}$, entonces $\frac{dw}{dv} = -\frac{1}{v^2}$, $0 < w < \infty \rightarrow \infty > \frac{1}{w} > 0$.

$$\begin{aligned}
\int_0^\infty \log(w) w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw &= \int_\infty^0 \log\left(\frac{1}{v}\right) v^{(a^{(k)}+1)} e^{-b^{(k)}v} \left(\frac{-1}{v^2}\right) dv \\
&= - \int_0^\infty \log(v) v^{(a^{(k)}-1)} e^{-b^{(k)}v} dv,
\end{aligned}$$

nuevamente haciendo un cambio de variable $u = \log(v) \rightarrow v = e^u$, entonces $\frac{dv}{du} = e^u$, $0 < v < \infty \rightarrow -\infty < u < \infty$.

$$\begin{aligned}
 &= - \int_{-\infty}^{\infty} u e^{u(a^{(k)}-1)} e^{-b^{(k)}e^u} e^u du, \\
 &= - \int_{-\infty}^{\infty} u e^{(a^{(k)}u - b^{(k)}e^u)} du
 \end{aligned}$$

NOTA: $\int_0^{\infty} w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw = \frac{\Gamma(a^{(k)})}{(b^{(k)})^{a^{(k)}}} = c(a^{(k)}, b^{(k)})$, ya que el integrando es el *kernel* de una $\mathcal{Iga}(a^{(k)}, b^{(k)})$. La expresión anterior es similar a la ecuación (B.1), por lo cual es válido hacer todos los cambios de variable que se aplicaron en dicha expresión. Entonces

$$\begin{aligned}
 c(a^{(k)}, b^{(k)}) &= \int_{\infty}^0 v^{(a^{(k)}+1)} e^{-b^{(k)}v} \left(\frac{-1}{v^2}\right) dv \\
 &= \int_0^{\infty} v^{(a^{(k)}-1)} e^{-b^{(k)}v} dv \\
 &= \int_{-\infty}^{\infty} e^{u(a^{(k)}-1)} e^{-b^{(k)}e^u} e^u du \\
 &= \int_{-\infty}^{\infty} e^{(a^{(k)}u - b^{(k)}e^u)} du \\
 \frac{\partial \log[c(a^{(k)}, b^{(k)})]}{\partial a^{(k)}} &= c(a^{(k)}, b^{(k)})^{-1} \int_{-\infty}^{\infty} u e^{(a^{(k)}u - b^{(k)}e^u)} du.
 \end{aligned}$$

Por lo tanto $\int_{-\infty}^{\infty} u e^{(a^{(k)}u - b^{(k)}e^u)} du = c(a^{(k)}, b^{(k)}) \frac{\partial \log[c(a^{(k)}, b^{(k)})]}{\partial a^{(k)}}$.

$$\begin{aligned}
\log[c(a^{(k)}, b^{(k)})] &= \log(\Gamma(a^{(k)})) - a^{(k)} \log(b^{(k)}) \\
\frac{\partial \log[c(a^{(k)}, b^{(k)})]}{\partial a^{(k)}} &= \psi(a^{(k)}) - \log(b^{(k)}) \\
\int_{-\infty}^{\infty} u e^{(a^{(k)}u - b^{(k)}e^u)} du &= \frac{\Gamma(a^{(k)})}{(b^{(k)})^{a^{(k)}}} \left\{ \psi(a^{(k)}) - \log(b^{(k)}) \right\} \\
\int_0^{\infty} \log(w) w^{-(a^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw &= \frac{\Gamma(a^{(k)})}{(b^{(k)})^{a^{(k)}}} \left\{ \log(b^{(k)}) - \psi(a^{(k)}) \right\}
\end{aligned}$$

Sustituyendo lo que valen $a^{(k)}$ y $b^{(k)}$

$$\begin{aligned}
&= \frac{\Gamma\left(\frac{\nu^{(k)}+2}{2}\right)}{\left(\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)})\xi_{t-1}^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)})\xi_t^2}{2}\right)^{\frac{\nu^{(k)}+2}{2}}} \\
&\times \left\{ \log\left(\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)})\xi_{t-1}^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)})\xi_t^2}{2}\right) - \psi\left(\frac{\nu^{(k)}+2}{2}\right) \right\} \\
&= \frac{\Gamma\left(\frac{\nu^{(k)}+2}{2}\right)}{\left(\frac{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)})\xi_{t-1}^2}{2}\right)^{\frac{\nu^{(k)}+2}{2}} \left[1 + \frac{(\xi_{t+1} - \phi^{(k)})\xi_t^2}{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)})\xi_{t-1}^2}\right]^{\frac{\nu^{(k)}+2}{2}}} \\
&\times \left\{ \log\left(\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)})\xi_{t-1}^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)})\xi_t^2}{2}\right) - \psi\left(\frac{\nu^{(k)}+2}{2}\right) \right\}
\end{aligned}$$

Otro cálculo pendiente en el *Paso-E* era resolver $\int_0^{\infty} w^{-(\alpha^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw$,
con

$$\alpha^{(k)} = \frac{\nu^{(k)}+2}{2} + 1, \quad b^{(k)} = \frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)})\xi_{t-1}^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)})\xi_t^2}{2}.$$

En la sección anterior vimos que el integrando es el kernel de una $\mathcal{Iga}(\alpha^{(k)}, b^{(k)})$
y por tanto

$$\begin{aligned}
\int_0^\infty w^{-(\alpha^{(k)}+1)} e^{-\frac{b^{(k)}}{w}} dw &= \frac{\Gamma(\alpha^{(k)})}{(b^{(k)})^{\alpha^{(k)}}} = \Gamma\left(\frac{\nu^{(k)}+2}{2}+1\right) \\
&\times \frac{1}{\left[\frac{\nu^{(k)}(\beta^{(k)})^2}{2} + \frac{(\xi_t - \phi^{(k)}\xi_{t-1})^2}{2} + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{2}\right]^{\frac{\nu^{(k)}+2}{2}+1}} \\
&= \frac{\left(\frac{\nu^{(k)}+2}{2}\right)\Gamma\left(\frac{\nu^{(k)}+2}{2}\right)2^{\frac{\nu^{(k)}+2}{2}+1}}{\left[\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2\right]^{\frac{\nu^{(k)}+2}{2}+1}} \\
&\times \frac{1}{\left[1 + \frac{(\xi_{t+1} - \phi^{(k)}\xi_t)^2}{\nu^{(k)}(\beta^{(k)})^2 + (\xi_t - \phi^{(k)}\xi_{t-1})^2}\right]^{\frac{\nu^{(k)}+2}{2}+1}}
\end{aligned}$$

Apéndice C

Códigos en MATLAB

En esta sección se presentan los códigos de los programas hechos en *MATLAB*, versión 7. que fueron usados a lo largo de esta tesis para los ejemplos y para el desarrollo del modelo *AR – ARCH* en general.

C.1. Códigos del Capítulo 2

- (a) Programa del ejemplo 2.4 usado para generar un gráfico de dispersión de una normal bivariada usando el algoritmo Gibbs Sampler.

```
function Gibbs_bivariate(n,m)

rho=.9;

y=[1;2];

sig=sqrt(1-rho^2);

xgibbs(1,:)= [10 10];

randn('state',sum(clock*100))

rand('state',sum(clock*100))
```

```

for i=2:n+m

    mu=y(1)+rho*(xgibbs(i-1,2)-y(2));

    xgibbs(i,1)= mu + sig*randn(1);

    mu=y(2) + rho*(xgibbs(i,1)-y(1));

    xgibbs(i,2)= mu + sig*randn(1);

end

close all

plot(xgibbs(m+1:n+m,1),...
xgibbs(m+1:n+m,2),'.','Color',[0 .5 1])

```

C.2. Códigos del Capítulo 3

- (a) Programa para simular un proceso ARMA(1, 1) de 200 observaciones.

```

%Simul_ARMA11

phi=-.45;

theta=1.2;

n=200;

WN=randn(n,1);

X(1)=randn();

for t=2:n

    X(t)=phi*X(t-1)+theta*WN(t-1)+WN(t);

```

```

end

close all

Y=X(2:n);

plot(Y)

```

(b) Programa para simular un proceso GARCH(1,1).

```

function r=GARCH11(sigma0,omega,beta,alpha,n,m)

randn('state',sum(clock*100))

e=randn([1 n+m]);

sigma=zeros(1,n+m);

y=zeros(1,n+m);

sigma(1)=sigma0;

y(1)=e(1)*sigma(1);

for i=2:n+m

    sigma(i)=sqrt(omega+alpha*y(i-1)^2+...
        beta*sigma(i-1)^2);

    y(i)=e(i)*sigma(i);

end

x=y(m+1:n+m);

vol=sigma(m+1:n+m);

```



```

r=x;

close all

subplot(2,1,1)

plot(x);

title('Retornos')

subplot(2,1,2)

plot(vol,'r')

title('Volatilidad')

whitebg('w')

```

(c) Programa para simular un proceso ARMA(1, 1)-GARCH(1, 1).

```

function ret=Sim_ARMA_GARCH(c,phi,theta,...
omega,beta,alpha,sigma0,n,m)

randn('state',sum(clock*100))

z=randn([1 n+m]);

sigma=zeros(1,n+m);

e=zeros(1,n+m);

x(1)=randn();

sigma(1)=sigma0;

e(1)=z(1)*sigma(1);

```

```

for t=2:n+m

    sigma(t)=sqrt(omega+alpha*e(t-1)^2+...
    beta*sigma(t-1)^2);

    e(t)=z(t)*sigma(t);

    x(t)=c+phi*x(t-1)+theta*e(t-1)+e(t);

end

r=x(m+1:n+m);

vol=sigma(m+1:n+m);

ret=r;

close all

subplot(2,1,1)

plot(r);

title('Retornos')

subplot(2,1,2)

plot(vol,'r')

title('Volatilidad')

```

C.3. Códigos del Capítulo 4

- (a) Programa *AR_ARCH.m*: Usado para la simulación de los datos de un proceso AR-ARCH de la manera planteada en el Capítulo 4.

```
function [r vol]=AR_ARCH(beta,nu,phi,n,m)
```

```

rand('state',sum(clock*100))
randn('state',sum(clock*100))

psi0=randn;
y(1)=(1/beta^2)*trnd(nu);
psi(1)=psi0*phi+y(1);
w(1)=(gamrnd((nu+1)/2,((psi(1)-phi*psi0)^2/2
+(nu*beta^2/2))^-1))^(-1);

for t=2:n+m
    y(t)=normrnd(0,sqrt(w(t-1)));
    psi(t)=phi*psi(t-1)+y(t);
    w(t)=(gamrnd((nu+1)/2,((psi(t)-phi*psi(t-1))^2/2
+(nu*beta^2/2))^-1))^(-1);
    sigma(t)=sqrt((nu*beta^2+y(t-1)^2)/(nu-1));
end

r=psi(m+1:n+m);
vol=sigma(m+1:n+m);

close all
subplot(2,1,1)
plot(r,'b')
xlim([1 n])
title('Retornos')
subplot(2,1,2)
plot(vol,'r')
xlim([1 n])
title('Volatilidad')

```

- (b) Programa *newtonQ.m*: Este programa implementa el *método de Newton* para resolver numéricamente el sistema

$$\nabla Q = \begin{pmatrix} \frac{\partial Q}{\partial \phi} \\ \frac{\partial Q}{\partial \beta} \\ \frac{\partial Q}{\partial \nu} \end{pmatrix} = \mathbf{0},$$

y así poder completar el paso-M en el algoritmo EM.

```
function y=newtonQ(x0,theta_k,x,tol);

n=length(x);
xt=x(3:n); %psi(t+1)
xt1=x(2:n-1); %psi(t)
xt2=x(1:n-2); %psi(t-1)
x00=x(1);
x1=x(2);
beta_k=theta_k(1);
nu_k=theta_k(2);
phi_k=theta_k(3);

Muk=((nu_k+2)./(nu_k*(beta_k)^2+(xt1-phi_k*xt2).^2
+(xt-phi_k*xt1).^2));
Nuk=log((1/2)*(nu_k*(beta_k)^2+(xt1-phi_k*xt2).^2
+(xt-phi_k*xt1).^2))-psi((1/2)*(nu_k+2));

fun_1 = inline('sum((Muk).*(xt-phi*xt1).*(xt1)
-((nu+1)/2).*(2./(nu*beta^2+(xt1-phi*xt2).^2))
.*(xt1-phi*xt2).*(xt2)+(Muk.*(xt1-phi*xt2))
.*(xt2))+((nu+1)/2)*(nu*beta^2/(nu*beta^2
+(x1-phi*x00)^2))*(2*(x1-x00*phi))/(nu*beta^2
*x00)');

fun_2 = inline('sum((nu+1)/2*(2./(nu*beta^2
+(xt1-phi*xt2).^2))*nu*beta-Muk*nu*beta)
-(1/beta)+(nu+1)/2*(nu*beta^2/(nu*beta^2+
(x1-phi*x00)^2))*(2*(x1-phi*x00)^2
```

```

/(nu*beta^3)');

fun_3=inline('sum((nu+1)/2*(beta^2./(nu*beta^2
+(xt1-phi*xt2).^2)))+(1/2)*log((nu*beta^2
+(xt1-phi*xt2).^2)/2)-(1/2)*psi((nu+1)/2)-(1/2)
*Nuk-(beta^2/2)*Muk)+(1/2)*psi((nu+1)/2)-(1/2)
*psi(nu/2)-1/(2*nu)-(1/2)*log(1+(x1-phi*x00)^2
/(nu*beta^2))+((nu+1)/2)*((nu*beta^2)
/(nu*beta^2+(x1-phi*x00)^2))
*((x1-phi*x00)^2/(nu^2*beta^2))');

%Valuación de la función

fx0_1 = feval(fun_1,Muk,x0(1),x0(2),x0(3),
x00,x1,xt,xt1,xt2);
fx0_2 = feval(fun_2,Muk,x0(1),x0(2),x0(3),
x00,x1,xt1,xt2);
fx0_3 = feval(fun_3,Muk,Nuk,x0(1),x0(2),x0(3)
,x00,x1,xt1,xt2);

h=10^-6;

while norm([fx0_1 fx0_2 fx0_3]) > tol

fx0_1 = feval(fun_1,Muk,x0(1),x0(2),x0(3)
,x00,x1,xt,xt1,xt2);
fx0_2 = feval(fun_2,Muk,x0(1),x0(2),x0(3)
,x00,x1,xt1,xt2);
fx0_3 = feval(fun_3,Muk,Nuk,x0(1),x0(2)
,x0(3),x00,x1,xt1,xt2);

%Cálculo de la Matriz Jacobiana, compuesta por
%las derivadas parciales evaluadas en el
%punto inicial.

Jac_f(1,1) = (feval(fun_1,Muk,x0(1)+h,
x0(2),x0(3),x00,x1,xt,xt1,xt2)-fx0_1)/h;

```

```

Jac_f(1,2) = (feval(fun_1,Muk,x0(1),
x0(2)+h,x0(3),x00,x1,xt,xt1,xt2)-fx0_1)/h;

Jac_f(1,3) = (feval(fun_1,Muk,x0(1),x0(2)
,x0(3)+h,x00,x1,xt,xt1,xt2)-fx0_1)/h;

Jac_f(2,1) = (feval(fun_2,Muk,x0(1)
+h,x0(2),x0(3),x00,x1,xt1,xt2)-fx0_2)/h;

Jac_f(2,2) = (feval(fun_2,Muk,x0(1)
,x0(2)+h,x0(3),x00,x1,xt1,xt2)-fx0_2)/h;

Jac_f(2,3) = (feval(fun_2,Muk,x0(1)
,x0(2),x0(3)+h,x00,x1,xt1,xt2)-fx0_2)/h;

Jac_f(3,1) = (feval(fun_3,Muk,Nuk,x0(1)
+h,x0(2),x0(3),x00,x1,xt1,xt2)-fx0_3)/h;

Jac_f(3,2) = (feval(fun_3,Muk,Nuk,x0(1)
,x0(2)+h,x0(3),x00,x1,xt1,xt2)-fx0_3)/h;

Jac_f(3,3) = (feval(fun_3,Muk,Nuk,x0(1)
,x0(2),x0(3)+h,x00,x1,xt1,xt2)-fx0_3)/h;

%Se resuelve el sistema

s=Jac_f\(-[fx0_1 fx0_2 fx0_3]');

%Se actualiza el paso

x0=x0+s';
end

y=x0;

```

- (c) Programa *EM_est.m*: Aquí se implementan de manera conjunta ambos pasos el *E* y el *M* para obtener la moda de la distribución *a posteriori*

de los parámetros.

```
function [iter params]=EM_est(x,theta0,tol)

max_iter=1200;
i=2;
theta(:,1)=theta0;
theta(:,2)=newtonQ(theta0,theta(:,1),x,tol);

while norm(theta(:,i)-theta(:,i-1))>tol
    if i==max_iter
        break;
    end
    i=i+1;
    theta(:,i)=newtonQ(theta(:,i-1)',theta(:,i-1),x,tol);
end

iter=i;

if i==max_iter
    disp('La solución no converge')
else
    fprintf('\n')
    fprintf('El valor de beta es %1.7f \n',theta(1,i))
    fprintf('El valor de nu es %1.7f \n',theta(2,i))
    fprintf('El valor de phi es %1.7f \n',theta(3,i))
    fprintf('\n')
end

params=theta(:,i);
```

- (d) Programa *AR_OLS.m*: Estima el modelo como si fuera un $AR(p)$, (con $p=1$ en el ejemplo del Capítulo 4) usando el método de mínimos cuadrados.

```
function [datos resid phi0]=AR_OLS(x,p,ind)
```

```

n=length(x);
fprintf('\n');

if ind==1
    x1=x-mean(x);
else
    x1=x;
end

for k=1:p
    for i=1:n-p
        A1(i,(p+1)-k)=x1(i+(k-1));
        Y(i)=x1(i+(k-1));
    end
end

for i=1:n-p
    Y(i)=x1(i+p);
end

phi0=((A1'*A1)^-1)*A1'*Y'

if ind==1
    Xt=A1*phi0+mean(x);
else
    Xt=A1*phi0;
end

for i=1:n-p
    y_est(i)=Xt(i);
end

y=x(p+1:n);
datos(1)=0;
datos(2:n)=y_est;

%Residuales

```



```

res=y-y_est;
res(1)=0;
res(2:n)=res(1:n-1);
resid=res;

```

- (e) Programa *maxver_tscaled.m*: Usa los residuales que se obtienen en la función *AR_OLS.m* y calcula los estimadores máximo verosímiles de los parámetros que definen a una distribución $\mathcal{St}(x|\mu, \beta, \nu)$ usando una función de la librería de *MATLAB Statistics Toolbox*.

```

function [beta0 nu0]=maxver_tscaled(x)

p=mle('tlocationsscale',x);
beta0=p(2)
nu0=p(3)

```

- (f) Programa *fitparams.m*: En este programa se hace una estimación preliminar de los parámetros para después usarlos como valores iniciales en el algoritmo EM.

```

function [beta nu phi]=fitparams(x,npred)

n=length(x);
x1=x(1:n-npred);
n1=length(x1);

%Estimaciones iniciales

[xs z1 phi0]=AR_OLS(x1,1,0); %xs = datos ajustados,
%y = residuales
z=z1(2:n1);
[beta0 nu0]=maxver_tscaled(z);

%Estimación usando el algoritmo E-M

```

```

[iter params]=EM_est(x1,[beta0 nu0 phi0],.00001);
beta=params(1);
nu=params(2);
phi=params(3);

```

- (g) Programa *fitmodel.m*: Con este programa se ajustan los datos con los parámetros estimados que se obtuvieron y también se hace predicción de observaciones futuras.

```

function [xp sig res]=fitmodel(x,s,beta,nu,phi,npred)

n=length(x);
x1=x(1:n-npred);
n1=length(x1);
xfit(1)=0;
xfit(2:n1)=phi.*x1(1:n1-1);
y=x1-xfit;

%Estimación de la volatilidad

sig(1)=sqrt((beta^2*nu)/(nu-1));
for t=2:n1
    sig(t)=sqrt((1/(nu-1))*y(t-1)^2+(beta^2*nu)/(nu-1));
end

volat=sig;
res=y./volat;

%Predicciones
%npred = # predicciones

for k=n1:n1+npred-1
    sig(k+1)=sqrt((1/(nu-1))*y(k)^2+(beta^2*nu)/(nu-1));
    y(k+1)=sqrt((1/(nu-1))*y(k)^2+(beta^2*nu)/(nu-1));
    x1(k+1)=phi*x1(k);
end

xp=cat(2,xfit,x1(n1+1:n1+npred));

```

```

close all
figure;
subplot(3,1,1)
plot(res,'color',[0 .5 1])
title('Residuales del proceso ARCH')
subplot(3,1,2)
plot(sig,'r')
title('Volatilidad')
subplot(3,1,3)
plot(x1)
title('Retornos')

figure;
hold on
plot(x,'g')
plot(xp,'r')
plot(xp(1:n1),'color',[0 .5 1])
xlim([1 n1+npred+10])
legend('Ret. originales','Ret. Pronosticados',
'Ret. ajustados',0)

figure;
hold on
plot(s,'b')
plot(sig,'r')
plot(sig(1:n1),'color',[1 .5 0])
xlim([1 n1+npred+10])
legend('Vol. originales','Vol Pronosticada',
'Vol. estimada',2)

fprintf('\n')
disp(' Retorno Real Predicción Vol. Real Predicción')
disp('-----')
for i=n1+1:n1+npred
disp(sprintf(' %1.6f %1.6f %1.6f %1.6f
',x(i),xp(i),s(i),sig(i)));
end

```

```
fprintf('\n')
```

- (h) Programa *ar_archfit.m*: Esta función permite comparar el método de estimación que construimos con el ajuste que hace *MATLAB* usando la librería de *GARCH Toolbox*, con la diferencia que se hace una modificación al reparametrizar los ajustes que se obtienen.

```
function ar_archfit(x,npred)
```

```
%Estas funciones son de la librería GARCH Toolbox
```

```
spec= garchset('Distribution','T','R',1,'Q', 1);  
pec = garchset(spec,'Display','off');  
[coef,error,llf,e,sigma_y] = garchfit(spec,x);  
S = garchfit(spec,x);
```

```
omega=garchget(coef,'K');  
alpha=garchget(coef,'ARCH');
```

```
%Reparametrización
```

```
nuu=(alpha^-1)+1;  
beta0=sqrt(omega*(nuu+1)/nuu)  
nu0=nuu  
phi0=garchget(coef,'AR')
```

```
[SigmaForecast,MeanForecast] = garchpred(S,x,npred)
```

- (i) Programa para simular trayectorias de precios utilizando el modelo de movimiento Browniano.

```
function SPaths=AssetPaths(S0,mu,sigma,T,NSteps,NPaths)
```

```
SPaths = zeros(NPaths,1+NSteps);
```

```
SPaths(:,1) = S0;
```

```

dt = T/NSteps;

mut = (mu-0.5*sigma^2)*dt;

sigt = sigma*sqrt(dt);

for i=1:NPaths

    for j=1: NSteps

        SPaths(i,j+1)=SPaths(i,j)*exp(mut + sigt*randn);

    end

end

```

- (j) Programa para simular precios suponiendo que los retornos se comportando como un modelo AR-ARCH; este programa es casi igual al programa *simAR_ARCH.m*, sólo que se considera que las primeras 2 observaciones son conocidas.

```

function x=simAR_ARCH2(xn1,xn2,beta,nu,phi,n)

tst=trnd(nu+1,1,n);

x=zeros(1,n);

y=zeros(1,n);

x0=xn2;

x(1)=xn1;

y(1)=x(1)-x0*phi;

for t=2:n

```

```
y(t)=sqrt((y(t-1)^2+nu*beta^2)/(1+nu))*tst(t);
```

```
x(t)=phi*x(t-1)+y(t);
```

```
end
```

- (k) Programa para valuar una opción tipo *call* considerando que el retorno del precio sigue un proceso AR-ARCH, utilizando el método de Monte Carlo.

```
function [sim_val,bls_price]=pricing_call(St,K,r,d,...  
vol_bls,beta,nu,phi,sim,xn1,xn2)
```

```
T=d/360;
```

```
R=zeros(sim,d);
```

```
S=zeros(sim,d+1);
```

```
rand('state',sum(clock*100))
```

```
randn('state',sum(clock*100))
```

```
for i=1:sim
```

```
    R(i,:)=simAR_ARCH2(xn1,xn2,beta,nu,phi,d);
```

```
    S(i,:)=ret2price(R(i,:)./100,St);
```

```
end
```

```
ST=S(:,d+1);
```

```
payoff=max(ST-K,0);
```

```
sim_val=mean(payoff)*exp(-r*T);
```

```
bls_price= bls(St,K,r,T,vol_bls);
```

- (l) Programa para valuar una opción tipo *call* con el modelo de Black-Scholes.

```
function Call= bls(S,K,r,tau,sigma,q)

if tau > 0

    d1 = (log(S/K) + ((r-q) +...
        0.5*sigma^2)*tau)/(sigma*sqrt(tau));

    d2 = d1 - sigma*sqrt(tau);

    N1 = 0.5*(1+erf(d1/sqrt(2)));
    N2 = 0.5*(1+erf(d2/sqrt(2)));

    Call = S*exp(-q*tau)*N1-K*exp(-r*tau)*N2;

else

    Call = max(S-K,0);

end
```

- (m) Programa para valuar una opción tipo *call* con el modelo de árboles binomiales.

```
function bin_price=binom(S,K,r,T,sigma,M)

dt = T/M;

u=exp(sigma*sqrt(dt));

d=1/u;

p = (exp(r*dt)-d)/(u-d);
```

```

q = 1-p;

W = max(S*u.^([M:-1:0]') .* d.^([0:M]') - K, 0);

for i = M:-1:1

    W = p*W(1:i) + q*W(2:i+1);

end

W = exp(-r*T)*W;

bin_price=W;

```


Bibliografía

- [1] BERGER, J. O. y BERNARDO, J. M. (1992), *On the development of reference priors*. Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, D. V. Lindley and A. F. M. Smith, eds). Oxford: Oxford University Press, 35-60 (with discussion).
- [2] BERNARDO, J. M. y SMITH, A. F. M. (1994), *Bayesian Theory*. Chichester, U.K. Wiley.
- [3] BERNARDO, J. M. (1980), *Bioestadística: Una Perspectiva Bayesiana*. Valencia, España. Vicens-Vives.
- [4] BOLLERSLEV, T. (1986), *A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*. Journal of Econometrics 31, 307-327.
- [5] BOLLERSLEV, T. (1987), *Generalized Autoregressive Conditional Heteroskedasticity*. The Review of Economics and Statistics, MIT Press, vol. 69(3), pages 542-47.
- [6] BROCKWELL, P. J. y DAVIS, R. A. (1987), *Time Series: Theory and Methods*. 2da Ed. Colorado, USA. Springer.
- [7] BROCKWELL, P. J. y DAVIS, R. A. (2002), *Introduction to Time Series and Forecasting*. 2da Ed. Colorado, USA. Springer.
- [8] CABALLERO, M. E., RIVERO, V. M., URIBE-BRAVO, G. y VELARDE, C. (2004), *Cadenas de Markov. Un enfoque elemental*. 1era Ed. México. Sociedad Matemática Mexicana.
- [9] CONTRERAS - CRISTÁN, A.(2008), *Notas de Licenciatura del Curso: Series de Tiempo*. Facultad de Ciencias, UNAM.

- [10] ENGLE, R. (1982), *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica*, 1982, vol. 50, issue 4, pages 987-1007.
- [11] ENGLE, R. y GRANGER, C. (1987), *Co-integration and Error Correction: Representation, Estimation, and Testing*. *Econometrica*, 1987, vol. 55, issue 2, pages 251-76 7.
- [12] FAN, J. y YAO, Q. (2003), *Nonlinear Time Series. Nonparametric and Parametric Methods*. Nueva York. Springer.
- [13] GAMERMAN, D., (1997), *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*. Londres. Chapman & Hall.
- [14] GLOSTEN, L. R., JAGANNATHAN y D.E. RUNKLE (1993), *On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on stocks*. *Journal of Finance*, 48, pp. 1,779-801.
- [15] GUTIÉRREZ-PEÑA, E., (1997), *Métodos Computacionales en la Inferencia Bayesiana*. 1era Ed., Vol. 6, No. 15. Instituto de Investigación en Matemáticas Aplicadas y Sistemas, UNAM.
- [16] HULL, J.C. y WHITE, A. (1987), *The Pricing of Options on Assets With Stochastic Volatilities*. 42, 281-300. *Journal of Finance*.
- [17] HULL, J.C. (2005), *Options, Futures and Other Derivatives*. 6ta Ed. Canadá. Prentice-Hall.
- [18] KANNAN, D. (1979), *An Introduction to Stochastic Processes*. North Holland series in probability and applied mathematics.
- [19] MARTINEZ, W. L. y MARTINEZ, A. R (2002), *Computational Statistics Handbook with MATLAB*. USA. Chapman & Hall/CRC.
- [20] NELSON, D. B. (1991), *Conditional heteroskedasticity in asset returns: A new approach*. *Econometrica* 59: 347-370.
- [21] PITT, M. K., CHATFIELD, C. y WALKER, S. G. (2002), *Constructing First-Order Stationary Autoregressive Models via Latent Processes*. 29, 657-663. *Scandinava Journal of Statistics*.

- [22] PITT, M. K. y WALKER, S. G. (2005), *Constructing Stationary Time Series Models Using Auxiliary Variables With Applications*. Vol. 100, No. 470, Theory and Methods. Journal of the American Statistical Association.
- [23] ROBERT, C. y CASELLA, G. (1999), *Monte Carlo Statistical Methods*. NY, USA. Springer.
- [24] SHUMWAY, R. H. y STOFFER, D. S. (2006), *Time Series Analysis and Its Applications With R Examples*. 2da Ed. USA. Springer.
- [25] STENTOFT L. (2004), *Pricing American Options when the Underlying Asset follows GARCH processes*. Dinamarca. School of Economics & Managment, University of Aarhus.
- [26] TANNER, M.A.(1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3rd ed. Springer, New York.
- [27] TSAY, R. S. (2002), *Analysis of Financial Time Series. Financial Econometrics*. Candá. Wiley & Sons.
- [28] WILMOTT, P. (2006), *Paul Wilmott on Quantitative Finance*. 2da Ed. Gran Bretaa. Wiley & Sons.
- [29] WÜRTZ, D., CHALABI, Y. y LUKSAN, L. (2006) *Parameter Estimation of ARMA Models with GARCH/APARCH Errors. An R and SPlus Software Implementation*. Volumen VV, artículo II. Journal of Statistical Software.