Construcción de un modelo AR-ARCH vía variables latentes

Miguel Angel Contreras Ortiz

Facultad de Ciencias, UNAM

Abril 2011

Introducción

- Pitt, Chatfield, Walker (2002) y Pitt, Walker(2005) desarrollaron artículos en los cuales exponen una metodología para construir series de tiempo estacionarias incorporando variables auxiliares; lo anterior con el fin de poder fijar una distribución marginal específica para las observaciones del proceso en estudio.

Introducción

- Pitt, Chatfield, Walker (2002) y Pitt, Walker(2005) desarrollaron artículos en los cuales exponen una metodología para construir series de tiempo estacionarias incorporando variables auxiliares; lo anterior con el fin de poder fijar una distribución marginal específica para las observaciones del proceso en estudio.
- Está metodología de incorporar variables auxiliares (ó latentes) en un modelo probabilístico tiene su motivación en técnicas de simulación estocástica usadas en la estadística Bayesiana como es el algoritmo Gibbs sampler.

Algoritmo Gibbs Sampler

- Para el caso de una distribución $f(\mathbf{x})$ definida en $\mathbb{R}^n$, el algoritmo Gibbs sampler permite simular muestras de una cadena de Makov $x^{(1)}, x^{(2)}, \ldots$ con distribución estacionaria $f(\mathbf{x})$.

Algoritmo Gibbs Sampler

- Para el caso de una distribución $f(\mathbf{x})$ definida en $\mathbb{R}^n$, el algoritmo Gibbs sampler permite simular muestras de una cadena de Makov $x^{(1)}, x^{(2)}, \ldots$ con distribución estacionaria $f(\mathbf{x})$.
- Una aplicación de este algoritmo se da cuando tenemos una densidad conjunta dada por $f(x, y_1, \ldots, y_n)$ y nos interesan características de la densidad marginal $f(x)$.

Algoritmo Gibbs Sampler

- Para el caso de una distribución $f(\mathbf{x})$ definida en $\mathbb{R}^n$, el algoritmo Gibbs sampler permite simular muestras de una cadena de Makov $x^{(1)}, x^{(2)}, \ldots$ con distribución estacionaria $f(\mathbf{x})$.
- Una aplicación de este algoritmo se da cuando tenemos una densidad conjunta dada por $f(x, y_1, \ldots, y_n)$ y nos interesan características de la densidad marginal $f(x)$.
- En este caso podemos usar el algoritmo Gibbs sampler para simular muestras $(x^{(i)}, y_1^{(i)}, \ldots, y_k^{(i)})$, $i = 1, 2, \ldots, n$ de $f(x, y_1, \ldots, y_k)$; entonces la muestra $\{x^{(i)}\}_{i=1}^{n}$ corresponderá a la densidad marginal $f(x)$.

Algoritmo Gibbs Sampler

- Para el caso de una distribución $f(\mathbf{x})$ definida en $\mathbb{R}^n$, el algoritmo Gibbs sampler permite simular muestras de una cadena de Makov $x^{(1)}, x^{(2)}, \ldots$ con distribución estacionaria $f(\mathbf{x})$.
- Una aplicación de este algoritmo se da cuando tenemos una densidad conjunta dada por $f(x, y_1, \ldots, y_n)$ y nos interesan características de la densidad marginal $f(x)$.
- En este caso podemos usar el algoritmo Gibbs sampler para simular muestras $(x^{(i)}, y_1^{(i)}, \ldots, y_k^{(i)})$, $i = 1, 2, \ldots, n$ de $f(x, y_1, \ldots, y_k)$; entonces la muestra $\{x^{(i)}\}_{i=1}^n$ corresponderá a la densidad marginal $f(x)$.
- Consideremos un caso donde la distribución conjunta es $f(X_1, X_2)$.

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.
- Las entradad del vector $\mathbf{X_t}$ son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Para la t-ésima iteración, dada la muestra $\mathbf{X_t} = (x_{t,1}, x_{t,2})$ hacemos los siguiente:

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.
- Las entradad del vector $\mathbf{X_t}$ son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Para la t-ésima iteración, dada la muestra $\mathbf{X_t} = (x_{t,1}, x_{t,2})$ hacemos los siguiente:
- 1. Generamos un valor $x_{t+1,1}$ de $f(X_1|X_2 = x_{t,2})$.

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.
- Las entradad del vector $\mathbf{X_t}$ son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Para la t-ésima iteración, dada la muestra $\mathbf{X_t} = (x_{t,1}, x_{t,2})$ hacemos los siguiente:
- 1. Generamos un valor $x_{t+1,1}$ de $f(X_1|X_2 = x_{t,2})$.
- 2. Generamos un valor $x_{t+1,2}$ de $f(X_2|X_1 = x_{t+1,1})$.

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.
- Las entradad del vector $\mathbf{X_t}$ son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Para la t-ésima iteración, dada la muestra $\mathbf{X_t} = (x_{t,1}, x_{t,2})$ hacemos los siguiente:
- 1. Generamos un valor $x_{t+1,1}$ de $f(X_1|X_2 = x_{t,2})$.
- 2. Generamos un valor $x_{t+1,2}$ de $f(X_2|X_1 = x_{t+1,1})$.
- 3. La t-ésima muestra de la cadena es $\mathbf{X_{t+1}} = (x_{t+1,1}, x_{t+1,2})$.

Algoritmo Gibbs Sampler

- Dado el vector inicial $\mathbf{X_0} = (x_{0,1}, x_{0,2})$, generamos muestras de $f(X_1, X_2)$ al muestrear $x_{1,1}$ de $f(X_1|X_2 = x_{0,2})$ y $x_{1,2}$ de $f(X_2|X_1 = x_{1,1})$.
- Las entradad del vector $\mathbf{X_t}$ son obtenidas alternadamente generando valores de ambas distribuciones condicionales. Para la t-ésima iteración, dada la muestra $\mathbf{X_t} = (x_{t,1}, x_{t,2})$ hacemos los siguiente:
- 1. Generamos un valor $x_{t+1,1}$ de $f(X_1|X_2 = x_{t,2})$.
- 2. Generamos un valor $x_{t+1,2}$ de $f(X_2|X_1 = x_{t+1,1})$.
- 3. La t-ésima muestra de la cadena es $\mathbf{X_{t+1}} = (x_{t+1,1}, x_{t+1,2})$.
- Al repetir los pasos 1, 2 y 3 un número n de veces grande, los valores $(x_{m+t,1}, x_{m+t,2})$ serán una muestra de $f(x_1, x_2)$ para $t = 1, 2, \ldots$ y $m \in \mathbb{Z}, 0 \leq m \leq n$.

Variables Latentes

- El método de variables latentes consiste en notar que en ocasiones es posible simular más facilmente 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 escencialmente cualquier variable aleatoria; a esta variable se le llama *variable auxiliar o variable latente*.

Variables Latentes

- El método de variables latentes consiste en notar que en ocasiones es posible simular más facilmente 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 escencialmente cualquier variable aleatoria; a esta variable se le llama *variable auxiliar o 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.

Variables Latentes

- Ejemplo: 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 $\mathbf{x} \sim \mathcal{N}b(\alpha, p)$.

Variables Latentes

- Ejemplo: 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 $\mathbf{x} \sim \mathcal{N}b(\alpha, p)$.
- En este ejemplo estamos considerando a λ como variable latente, ya que

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

Variables Latentes

- Ejemplo: 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 $\mathbf{x} \sim \mathcal{N}b(\alpha, p)$.
- En este ejemplo estamos considerando a λ como variable latente, ya que

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

- Mas aún, si se conocieran las distribuciones condicionales completas, es decir $f(\mathbf{x}|\lambda)$ y $f(\lambda|\mathbf{x})$, se puede realizar la simulación usando el algoritmo Gibbs Sampler.

Inferencia Bayesiana

- En el enfoque Bayesiano de la Estadística, la incertidumbre presente en un modelo dado, $p(\mathbf{x}|\theta)$, es representado a través de una *distribución inicial o a priori* $p(\theta)$ sobre los posibles valores del parámetro desconocido que define al modelo.

Inferencia Bayesiana

- En el enfoque Bayesiano de la Estadística, la incertidumbre presente en un modelo dado, $p(\mathbf{x}|\theta)$, es representado a través de una *distribución inicial o a priori* $p(\theta)$ sobre los posibles valores del parámetro desconocido que define al modelo.
- El Teorema de Bayes,

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

permite entonces incorporar la información contenida en un conjunto de datos $\mathbf{x} = (x_1, \ldots, x_n)$, produciendo una descripción conjunta de la incertidumbre sobre los valores de los parámetros del modelo a través de la *distribución final o posterior* $p(\theta|\mathbf{x})$.

Inferencia Bayesiana

- 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" o bien "igual salvo por un factor que no depende de θ".

Inferencia Bayesiana

- 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" o bien "igual salvo por un factor que no depende de θ".

- 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 final.

Algoritmo E-M

- El algoritmo EM (Expectation-Maximization) es un procedimiento iterativo para calcular la moda de la distribución final o los estimadores máximo verosimiles cuando se tienen datos incompletos.

Algoritmo E-M

- El algoritmo EM (Expectation-Maximization) es un procedimiento iterativo para calcular la moda de la distribución final o los estimadores máximo verosimiles cuando se tienen datos incompletos.
- En la formulación usual del algoritmo EM, el vector "completo" de datos **x** es conformado tanto por "datos observados" **y** como por "datos faltantes" **w**, es decir que los elementos de **w** se pueden considerar como variables latentes.

Algoritmo E-M

- El algoritmo EM (Expectation-Maximization) es un procedimiento iterativo para calcular la moda de la distribución final o los estimadores máximo verosimiles cuando se tienen datos incompletos.
- En la formulación usual del algoritmo EM, el vector "completo" de datos **x** es conformado tanto por "datos observados" **y** como por "datos faltantes" **w**, es decir que los elementos de **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.

Algoritmo E-M

- Cada iteración del algoritmo EM consiste de dos pasos: el paso **E** (paso de la esperanza) y el paso **M** (paso de la maximización).

Algoritmo E-M

- 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 final de nuestro interés $p(\theta|\mathbf{y})$ y sea $p(\theta|\mathbf{y}, \mathbf{w})$ la distribución final aumentada.

Algoritmo E-M

- 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 final de nuestro interés $p(\theta|\mathbf{y})$ y sea $p(\theta|\mathbf{y}, \mathbf{w})$ la distribución final aumentada.
- Por último definamos a $p(\mathbf{w}|\mathbf{y}, \theta^{(k)})$ como la distribución condicional de los datos latentes $\mathbf{w}$ dado el valor k-ésimo de la moda de la distribución final y dados los datos observados; entonces los pasos E y M en la $(k+1)$-ésima iteración son:

Algoritmo E-M

- **paso** $-$ **E**. Calcular $Q(\theta, \theta^{(k)}) = \mathbb{E}[\log p(\theta|\mathbf{y}, \mathbf{w})]$ con respecto a $p(\mathbf{w}|\mathbf{y}, \theta^{(k)})$, es decir,

$$Q(\theta, \theta^{(k)}) = \int_W \log p(\theta|\mathbf{y}, \mathbf{w}) p(\mathbf{w}|\mathbf{y}, \theta^{(k)}) d\mathbf{w}$$

y

Algoritmo E-M

- **paso** $-$ **E**. Calcular $Q(\theta, \theta^{(k)}) = \mathbb{E}[\log p(\theta|\mathbf{y}, \mathbf{w})]$ con respecto a $p(\mathbf{w}|\mathbf{y}, \theta^{(k)})$, es decir,

$$Q(\theta, \theta^{(k)}) = \int_W \log p(\theta|\mathbf{y}, \mathbf{w}) p(\mathbf{w}|\mathbf{y}, \theta^{(k)}) d\mathbf{w}$$

y

- **paso** $-$ **M**. Maximizar $Q(\theta, \theta^{(k)})$ con respecto a θ. Se toma

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta, \theta^{(k)}).$$

Algoritmo E-M

- **paso** $-$ **E**. Calcular $Q(\theta, \theta^{(k)}) = \mathbb{E}[\log p(\theta|\mathbf{y}, \mathbf{w})]$ con respecto a $p(\mathbf{w}|\mathbf{y}, \theta^{(k)})$, es decir,

$$Q(\theta, \theta^{(k)}) = \int_W \log p(\theta|\mathbf{y}, \mathbf{w}) p(\mathbf{w}|\mathbf{y}, \theta^{(k)}) d\mathbf{w}$$

  y

- **paso** $-$ **M**. Maximizar $Q(\theta, \theta^{(k)})$ con respecto a θ. Se toma

$$\theta^{(k+1)} = \arg \max_{\theta} Q(\theta, \theta^{(k)}).$$

- Después de ésto regresamos al paso-E. El algoritmo se itera hasta que $|Q(\theta^{(k+1)}, \theta^{(k)}) - Q(\theta^{(k)}, \theta^{(k)})|$ o bien $||\theta^{(k+1)} - \theta^{(k)}||$ sea suficientemente pequeño.

Series de tiempo

- 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.

Series de tiempo

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

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

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

Series de tiempo estacionarias

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

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

Series de tiempo estacionarias

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

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

- En otras palabras, toda la estructura probabilística de un proceso completamente estacionario no cambia al correr en el tiempo

Series de tiempo estacionarias

- Un proceso de 2^o orden $\{X_t\}_{t\in T}$ se llama *débilmente estacionario o estacionario de* 2^o *orden* si para cada $n \in \{1, 2, \ldots\}$, para cualesquiera $t_1, t_2, \ldots, t_n \in T$ y para todo τ tal que $t_1 + \tau, \ldots, t_n + \tau \in T$, todos los momentos de orden 1 y 2 del vector $(X_{t_1}, \ldots, X_{t_n})$ son iguales a los correspondientes momentos de orden 1 y 2 del vector $(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$.

Series de tiempo estacionarias

- Un proceso de 2^o orden $\{X_t\}_{t\in T}$ se llama *débilmente estacionario o estacionario de* 2^o *orden* si para cada $n \in \{1, 2, \ldots\}$, para cualesquiera $t_1, t_2, \ldots, t_n \in T$ y para todo τ tal que $t_1 + \tau, \ldots, t_n + \tau \in T$, todos los momentos de orden 1 y 2 del vector $(X_{t_1}, \ldots, X_{t_n})$ son iguales a los correspondientes momentos de orden 1 y 2 del vector $(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau})$.
- Algunas implicaciones de esta definición son

$$\mathbb{E}(X_t) = \mu, \quad \forall t$$

$$\text{Cov}(X_t, X_{t+\tau}) = \gamma_\tau, \quad \forall t.$$

Series de tiempo estacionarias

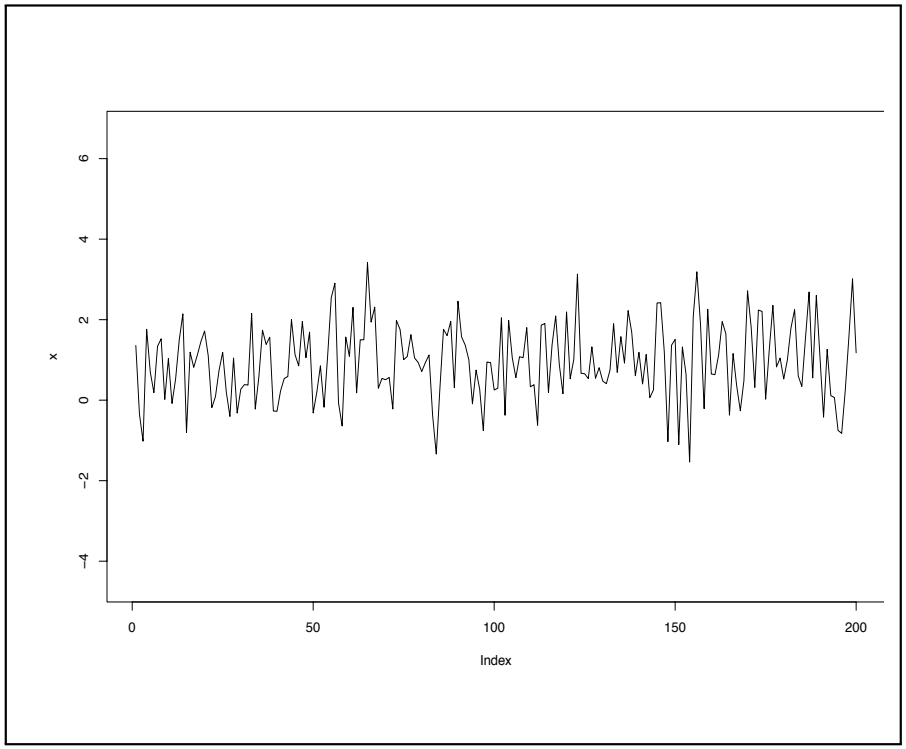
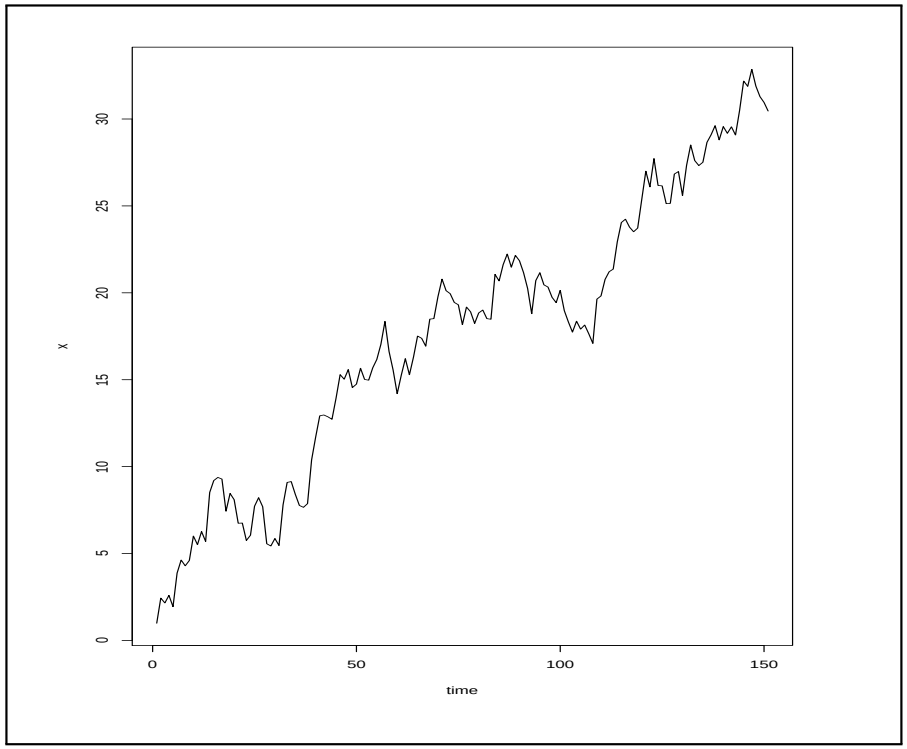


Figura: (a) Proceso estacionario



(b) Proceso no-estacionario

Modelos 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 $t \in \{0, \pm 1, \pm 2, \ldots\}$ se tiene

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_t, \quad (1)$$

donde $\{\varepsilon_t\}$ es una sucesión de v.a's no-correlacionadas, tal que $\forall t$, $\varepsilon_t \sim N(0, \sigma_\varepsilon)$.

Modelos ARMA(p,q)

- La ecuación (1) se puede escribir usando la notación de operadores de retraso $BX_t = X_{t-1}$ como

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

con

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

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

Modelos ARMA(p,q)

- La ecuación (1) se puede escribir usando la notación de operadores de retraso $BX_t = X_{t-1}$ como

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

con

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

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

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

Modelos ARMA(p,q)

- La ecuación (1) se puede escribir usando la notación de operadores de retraso $BX_t = X_{t-1}$ como

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

con

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

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

- **(a)** Si $\phi_p(z) \equiv 1$ entonces $X_t = \theta(B)\varepsilon_t$ se reduce a un modelo MA(q).
- **(b)** Si $\theta_q(z) \equiv 1$ entonces $\phi(B)X_t = \varepsilon_t$ se reduce un modelo AR(p).

Modelos ARMA(p,q)

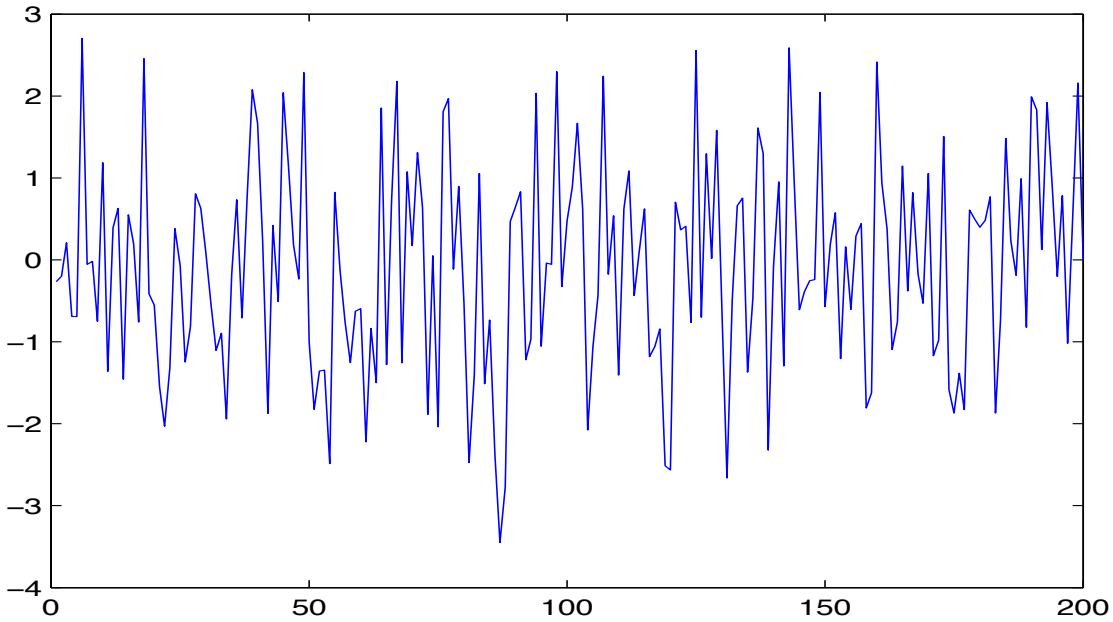


Figura: Ejemplo de un proceso ARMA(1,1).

Modelos ARCH(q)

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

$$Y_t = \sigma_t \eta_t, \quad \text{donde} \quad \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_q Y_{t-q}^2,$$

para constantes $\alpha_0 \geq 0, \alpha_j \geq 0$ y $\{\eta_t\}_t$ v.a.'s i.i.d. con alguna distribución paramétrica $\mathcal{D}_\vartheta$, con media 0 y varianza es 1 (comúnmente se usa Normal o t-Student).

Modelos ARCH(q)

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

$$Y_t = \sigma_t \eta_t, \quad \text{donde} \quad \sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_q Y_{t-q}^2,$$

  para constantes $\alpha_0 \geq 0, \alpha_j \geq 0$ y $\{\eta_t\}_t$ v.a.'s i.i.d. con alguna distribución paramétrica $\mathcal{D}_\vartheta$, con media 0 y varianza es 1 (comúnmente se usa Normal o t-Student).
- 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.

Modelos ARCH(q)

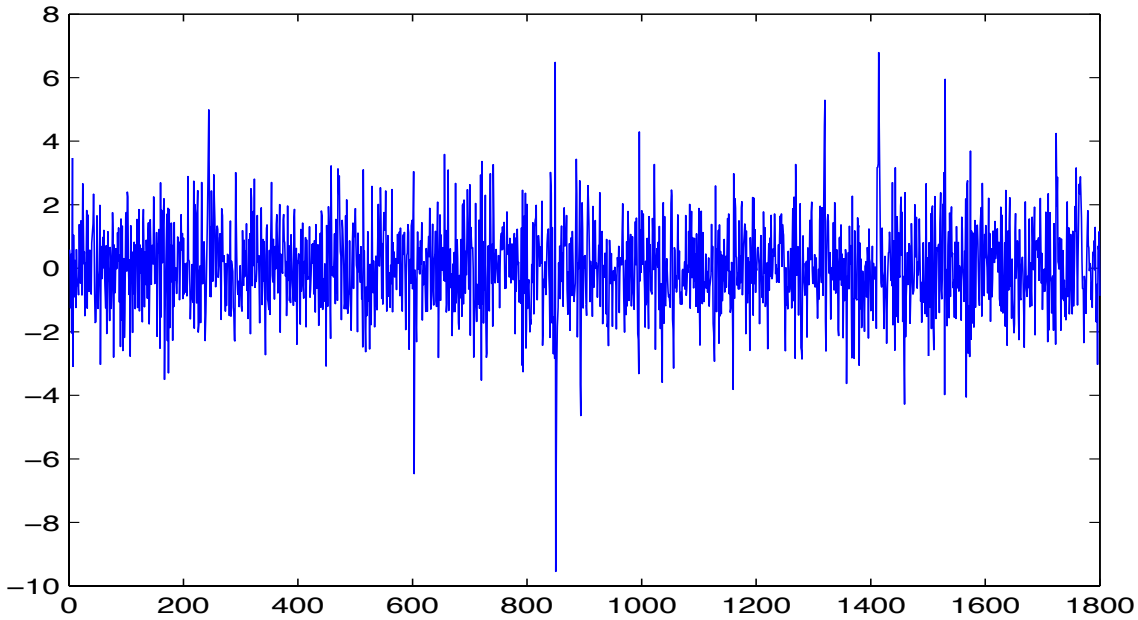


Figura: Ejemplo de un proceso ARCH(1).

Modelo ARCH(1)

- Dentro de este tipo de modelos el ARCH(1) es el más utilizado para modelar la voatilidad de series financieras

$$\sigma_t^2(Y_t|Y_{t-1}) = \alpha_0 + \alpha_1 Y_{t-1}^2,$$

donde para una serie de precios $S_1, S_2, \ldots$ se toma $Y_t = \log(S_t/S_{t-1})$ (conocidos en el ámbito financiero como *log-retornos* o simplemente *retornos*).

Modelo GARCH(q, p)

- El modelo ARCH ha sido extendido en varias direcciones. La más importante de ésas es la extensión para incluir un componente autorregresivo en la varianza condicional; esta extensión se conoce como *modelo ARCH generalizado (GARCH)* debido a Bollerslev (1986) y Taylor (1986).

Modelo GARCH(q, p)

- El modelo ARCH ha sido extendido en varias direcciones. La más importante de ésas es la extensión para incluir un componente autorregresivo en la varianza condicional; esta extensión se conoce como *modelo ARCH generalizado (GARCH)* debido a Bollerslev (1986) y Taylor (1986).
- El proceso $\{Y_t\}_t$ sigue un proceso GARCH(p, q) si

$$Y_t = \sigma_t \eta_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \qquad (2)$$

donde nuevamente $\{\eta_t\}$ 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 $\sum_{i=1}^{\text{máx}(p,q)} (\alpha_i + \beta_i) < 1$.

Modelo GARCH(q, p)

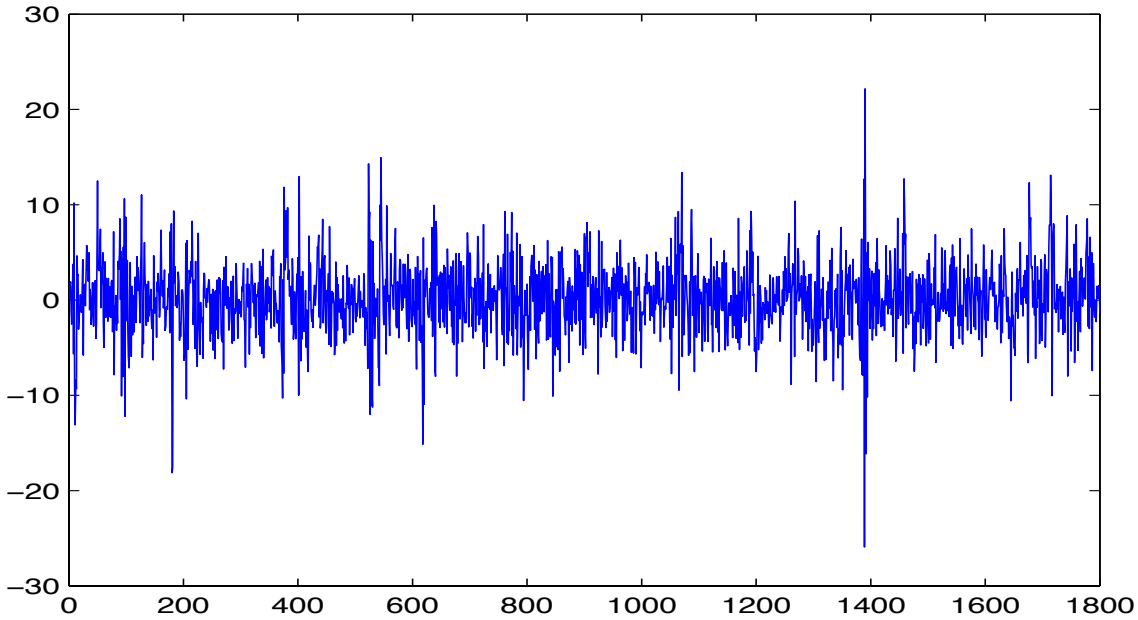


Figura: Realización de un proceso GARCH(1,1).

Modelo ARMA(m, n)-GARCH(q, p)

- 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

Modelo ARMA(m, n)-GARCH(q, p)

- 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
-

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

Modelo ARMA(m, n)-GARCH(q, p)

- 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
- $$X_t = \sum_{k=1}^{m} \phi_k X_{t-k} + Y_t + \sum_{j=1}^{n} \theta_j Y_{t-j},$$
- $$Y_t = \sigma_t \eta_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

donde η_t es i.i.d. de acuerdo a una cierta distribución con media cero y varianza unitaria.

Los modelos de Pitt y Walker

- Consideremos una sucesión de variables aleatorias

$$Y_1 \to W_1 \to Y_2 \to W_2 \to \cdots \to W_{n-1} \to Y_n. \qquad (3)$$

Los modelos de Pitt y Walker

- Consideremos una sucesión de variables aleatorias

$$Y_1 \to W_1 \to Y_2 \to W_2 \to \cdots \to W_{n-1} \to Y_n. \qquad (3)$$

- El proceso (3) 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 $t = 3, 4, 5, \ldots$

Los modelos de Pitt y Walker

- Consideremos una sucesión de variables aleatorias

$$Y_1 \to W_1 \to Y_2 \to W_2 \to \cdots \to W_{n-1} \to Y_n. \qquad (3)$$

- El proceso (3) 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 $t = 3, 4, 5, \ldots$
- De acuerdo a la estructura en que evoluciona el proceso hay una clara conexión con el Gibbs sampler.

Los modelos de Pitt y Walker

- 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 $f_{W_t|Y_t}(w|y)(= f_{W|Y}(w|y))$.

Los modelos de Pitt y Walker

- 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 $f_{W_t|Y_t}(w|y)(= f_{W|Y}(w|y))$.
- Pitt, Chatfield, Walker (2002) y Pitt, Walker (2005) proponen modelos de series de tiempo usando el esquema (3) donde el proceso $\{Y_1, Y_2, \ldots\}$ corresponde a las observaciones de algún fenómeno en la naturaleza y el proceso $\{W_1, W_2, \ldots\}$ corresponde a variables latentes que se introducen con el fin de especificar la dependencia estocástica entre Y_{t+1} y Y_t.

Los modelos de Pitt y Walker

- Dadas las densidades $f_{W_t|Y_t}$ y $f_{Y_{t+1}|W_t}$ tenemos que

$$f_{Y_{t+1}|Y_t}(y|y_t) = \int_{-\infty}^{\infty} f_{Y_{t+1}|W_t}(y|w) f_{W_t|Y_t}(w|y_t) dw. \qquad (4)$$

Los modelos de Pitt y Walker

- Dadas las densidades $f_{W_t|Y_t}$ y $f_{Y_{t+1}|W_t}$ tenemos que

$$f_{Y_{t+1}|Y_t}(y|y_t) = \int_{-\infty}^{\infty} f_{Y_{t+1}|W_t}(y|w) f_{W_t|Y_t}(w|y_t) dw. \qquad (4)$$

- Por construcción, el proceso $\{Y_1, W_1, Y_2, W_2, \ldots\}$ 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.

Modelo ARCH vía variables latentes

- Para construir un modelo tipo ARCH estrictamente estacionario, Pitt y Walker (2005) proponen un esquema como en (3) 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 $f_{Y_t}(y) = \mathcal{S}t(y|0, \frac{1}{\beta^2}, \nu)$.

Modelo ARCH vía variables latentes

- Para construir un modelo tipo ARCH estrictamente estacionario, Pitt y Walker (2005) proponen un esquema como en (3) 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 $f_{Y_t}(y) = \mathcal{S}t(y|0, \frac{1}{\beta^2}, \nu)$.
- 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 $Y_{t+1}|Y_t \sim \mathcal{S}t(y_{t+1}|0, \lambda = \frac{\nu+1}{\nu\beta^2+(y_t)^2}, \alpha = \nu + 1)$.

Modelo ARCH vía variables latentes

- Para construir un modelo tipo ARCH estrictamente estacionario, Pitt y Walker (2005) proponen un esquema como en (3) 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 $f_{Y_t}(y) = \mathcal{S}t(y|0, \frac{1}{\beta^2}, \nu)$.
- 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 $Y_{t+1}|Y_t \sim \mathcal{S}t(y_{t+1}|0, \lambda = \frac{\nu+1}{\nu\beta^2+(y_t)^2}, \alpha = \nu + 1)$.
- La varianza condicional tiene la siguiente forma

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

Modelo AR-ARCH

- El proceso $\{Y_t\}_t$ jugará el papel de las innovaciones o ruido en un modelo AR(1) para las observaciones $\{X_t\}_t$.

Modelo AR-ARCH

- El proceso $\{Y_t\}_t$ jugará el papel de las innovaciones o ruido en un modelo AR(1) para las observaciones $\{X_t\}_t$.
- Consideremos la ecuación que define a un proceso AR(1)

$$X_t = \phi X_{t-1} + Y_t, \tag{5}$$

donde $|\phi| < 1$ para garantizar que el proceso sea estacionario.

Modelo AR-ARCH

- Podemos pensar en (5) como una transformación entre dos vectores aleatorios

$$T(Y_1, \ldots, Y_n) = (X_1, \ldots, X_n),$$

donde dado un valor inicial constante $X_0 = x_0$ para $\{X_t\}_t$, cada entrada de $(X_1, \ldots, X_n)$ está definida por (5).

Modelo AR-ARCH

- Podemos pensar en (5) como una transformación entre dos vectores aleatorios

$$T(Y_1, \ldots, Y_n) = (X_1, \ldots, X_n),$$

donde dado un valor inicial constante $X_0 = x_0$ para $\{X_t\}_t$, cada entrada de $(X_1, \ldots, X_n)$ está definida por (5).

- La transformación inversa

$$T^{-1}(X_1, \ldots, X_n) = (Y_1, \ldots, Y_n)$$

queda definida por

$$Y_t = X_t - \phi X_{t-1}.$$

Modelo AR-ARCH

- Así entonces la función de verosimilitud para las observaciones $x_1, \ldots, x_n$ del proceso $\{X_n\}_n$ está dada por

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{Y_1,\ldots,Y_n}(x_1-\phi x_0,\ldots,x_n-\phi x_{n-1})\times|J_{T^{-1}}|$$

$$= f_{Y_1}(x_1-\phi x_0)\times\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1}),$$

Modelo AR-ARCH

- Así entonces la función de verosimilitud para las observaciones $x_1, \ldots, x_n$ del proceso $\{X_n\}_n$ está dada por

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{Y_1,\ldots,Y_n}(x_1-\phi x_0,\ldots,x_n-\phi x_{n-1})\times|J_{T^{-1}}|$$

$$= f_{Y_1}(x_1-\phi x_0)\times\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1}),$$

- donde $f_{Y_1}(x_1-\phi x_0) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(\frac{1}{\nu\beta^2\pi}\right)^{\frac{1}{2}}\frac{1}{[1+\frac{(x_1-\phi x_0)^2}{\nu\beta^2}]^{\frac{\nu+1}{2}}}$
  la cual es una distribución $\mathcal{S}t(x_1|\phi x_0,\lambda,\nu)$ con $\lambda = \frac{1}{\beta^2}$.

Modelo AR-ARCH

- Por otra parte, $f_{Y_{t+1}|Y_t}(x_{t+1} - \phi x_t | x_t - \phi x_{t-1})$

$$= \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})} \left(\frac{1}{(\nu\beta^2 + (x_t - \phi x_{t-1})^2)\pi} \right)^{\frac{1}{2}} \frac{1}{[1 + \frac{(x_{t+1} - \phi x_t)^2}{\nu\beta^2 + (x_t - \phi x_{t-1})^2}]^{\frac{\nu+2}{2}}},$$

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

Modelo AR-ARCH

- Por otra parte, $f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})$

$$= \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})}\left(\frac{1}{(\nu\beta^2+(x_t-\phi x_{t-1})^2)\pi}\right)^{\frac{1}{2}}\frac{1}{[1+\frac{(x_{t+1}-\phi x_t)^2}{\nu\beta^2+(x_t-\phi x_{t-1})^2}]^{\frac{\nu+2}{2}}},$$

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

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

Estimación de parámetros vía el algoritmo EM

- Como se dijo anteriormente, el objetivo del algoritmo EM es encontrar la moda de la distribución final y, recordando que si la distribución final es proporcional a la verosimilitud, al encontrar su moda estaríamos encontrando el estimador o los estimadores máximo verosímiles.

Estimación de parámetros vía el algoritmo EM

- Como se dijo anteriormente, el objetivo del algoritmo EM es encontrar la moda de la distribución final y, recordando que si la distribución final es proporcional a la verosimilitud, al encontrar su moda estaríamos encontrando el estimador o los estimadores máximo verosímiles.
- La expresión de la verosimilitud aumentada considerando el conjunto de datos completos sería la siguiente

$$f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}}|\nu,\beta,\phi(x_1,\ldots,x_n,w_1\ldots,w_{n-1})$$

$$=\prod_{t=1}^{n-1}\left\{f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t)\times f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1})\times f_{Y_1}(x_1-\phi x_0)\right\}$$

Estimación de parḿetros vía el algoritmo EM

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

$$\begin{aligned} & \pi(\nu, \beta, \phi | x_1, \ldots, x_n, w_1, \ldots, w_{n-1}) \\ \propto \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi \times \pi(\nu, \beta, \phi) \\ = \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi. \end{aligned}$$

Estimación de parḿetros vía el algoritmo EM

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

$$\begin{aligned} & \pi(\nu, \beta, \phi | x_1, \ldots, x_n, w_1, \ldots, w_{n-1}) \\ \propto \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi \times \pi(\nu, \beta, \phi) \\ = \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi. \end{aligned}$$

- Se pretende usar el algoritmo EM para encontrar a $\widehat{\nu}, \widehat{\beta}$ y $\widehat{\phi}$.

Estimación de parmetros vía el algoritmo EM

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

$$\begin{aligned} & \pi(\nu, \beta, \phi | x_1, \ldots, x_n, w_1, \ldots, w_{n-1}) \\ \propto \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi \times \pi(\nu, \beta, \phi) \\ = \quad & f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}} | \nu, \beta, \phi. \end{aligned}$$

- Se pretende usar el algoritmo EM para encontrar a $\widehat{\nu}, \widehat{\beta}$ y $\widehat{\phi}$.
- La distribución condicional de las variables latentes dados los datos observados y los parámetros será obtener:

Estimación de parámetros vía el algoritmo EM

- $$\frac{f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}|\nu,\beta,\phi}}{f_{X_1,\ldots,X_n|\nu,\beta,\phi}} = \frac{\prod_{t=1}^{n-1}\left\{ f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t) f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1}) \right\}}{\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})}$$

Estimación de parámetros vía el algoritmo EM

- $$\frac{f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}|\nu,\beta,\phi}}{f_{X_1,\ldots,X_n|\nu,\beta,\phi}} = \frac{\prod_{t=1}^{n-1}\left\{ f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t) f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1}) \right\}}{\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})}$$
- $$f_{W_1,\ldots,W_{n-1}|X_1,\ldots,X_n,\nu,\beta,\phi} = \prod_{t=1}^{n-1}\left\{ \frac{f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t) f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1})}{f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})} \right\}.$$

Estimación de parámetros vía el algoritmo EM

- $$\frac{f_{X_1,\ldots,X_n,W_1,\ldots,W_{n-1}|\nu,\beta,\phi}}{f_{X_1,\ldots,X_n|\nu,\beta,\phi}} = \frac{\prod_{t=1}^{n-1}\left\{ f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t) f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1}) \right\}}{\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})}$$
- $$f_{W_1,\ldots,W_{n-1}|X_1,\ldots,X_n,\nu,\beta,\phi} = \prod_{t=1}^{n-1}\left\{ \frac{f_{Y_{t+1}|W_t}(x_{t+1}-\phi x_t|w_t) f_{W_t|Y_t}(w_t|x_t-\phi x_{t-1})}{f_{Y_{t+1}|Y_t}(x_{t+1}-\phi x_t|x_t-\phi x_{t-1})} \right\}.$$
- Posteriormente se toma el logaritmo de la distribución final resultante y se procede a implementar los pasos E y M.

Ajuste de datos simulados

- 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 las observaciones $\{X_t\}_t$ con errores que siguen un proceso ARCH(1) como plantea el artículo de Pitt y Walker (2005).

Ajuste de datos simulados

- 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 las observaciones $\{X_t\}_t$ con errores que siguen un proceso ARCH(1) como plantea el artículo de Pitt y Walker (2005).
- Se ajustó el modelo estimando los parámetros vía el algoritmo EM. Para verificar la efectividad de predicción en el modelo, se realizaron 5 predicciones de la volatilidad de los datos y se compararon con las observaciones reales.

Ajuste de datos simulados

- Los verdaderos valores de los parámetros para los datos simulados fueron:

$$(\beta, \nu, \phi) = \ (2.5,4.5,0.95).$$

Ajuste de datos simulados

- Los verdaderos valores de los parámetros para los datos simulados fueron:

$$(\beta, \nu, \phi) = \ (2.5,4.5,0.95).$$

- Las estimaciones finales de los parámetros después de implementar el algoritmo EM fueron las siguientes:

$$(\hat{\beta}, \hat{\nu}, \hat{\phi}) = \ (2.5036478,\ 4.4077945,\ 0.9485942).$$

Ajuste de datos simulados

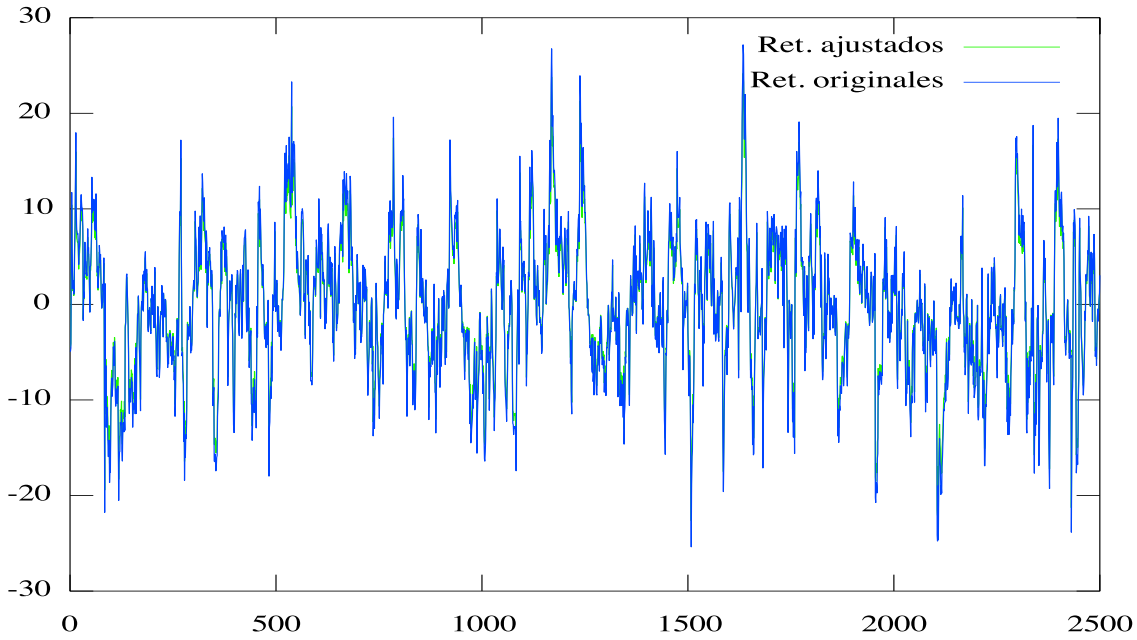


Figura: Datos estimados y originales.

Ajuste de datos simulados

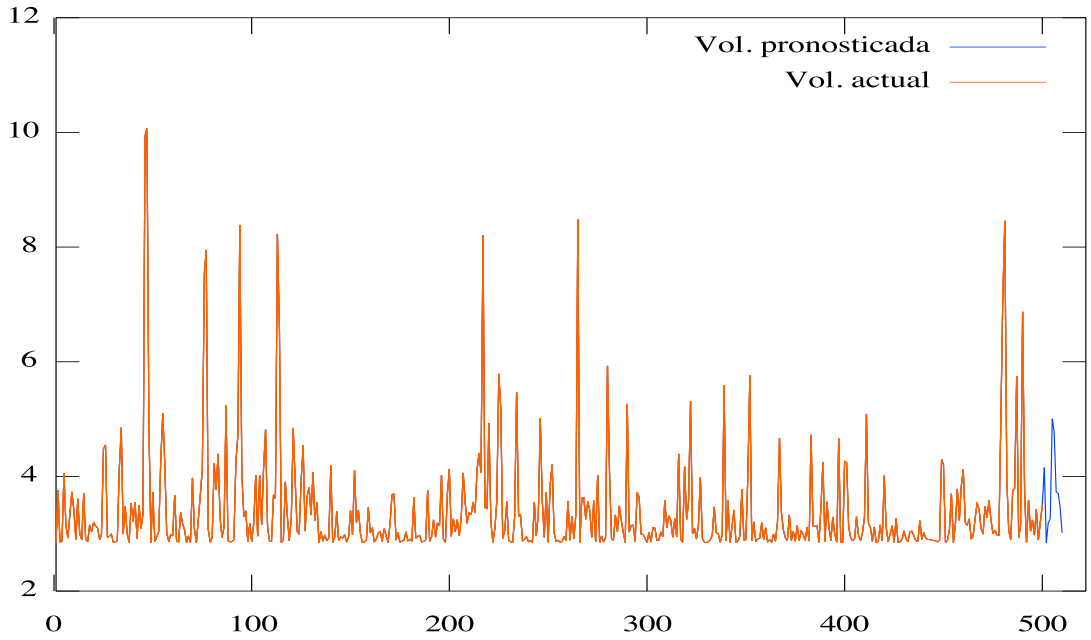


Figura: Volatilidad estimada y real.

Ajuste de datos reales

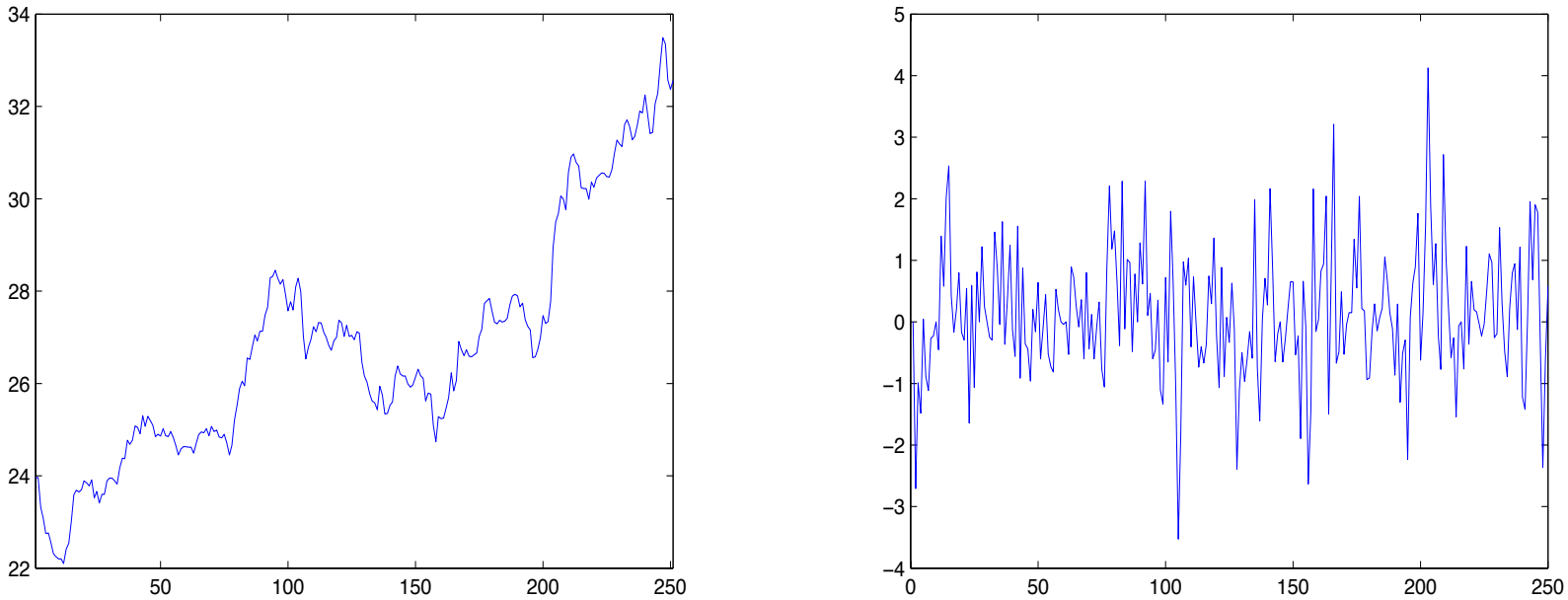


Figura: Precios y retornos de la acción de AT&T.

Ajuste de datos reales

- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).

Ajuste de datos reales

- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).
- Se ajustó la serie con el modelo propuesto y se comparó contra 3 modelos alternativos:

Ajuste de datos reales

- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).
- Se ajustó la serie con el modelo propuesto y se comparó contra 3 modelos alternativos:
- Un AR(1) con errores distribuidos $N(0, \sigma^2)$.

Ajuste de datos reales

- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).
- Se ajustó la serie con el modelo propuesto y se comparó contra 3 modelos alternativos:
- Un AR(1) con errores distribuidos $N(0, \sigma^2)$.
- Un MA(1)-ARCH(1) con errores distribuidos $\mathcal{S}t(0, 1, \nu)$.

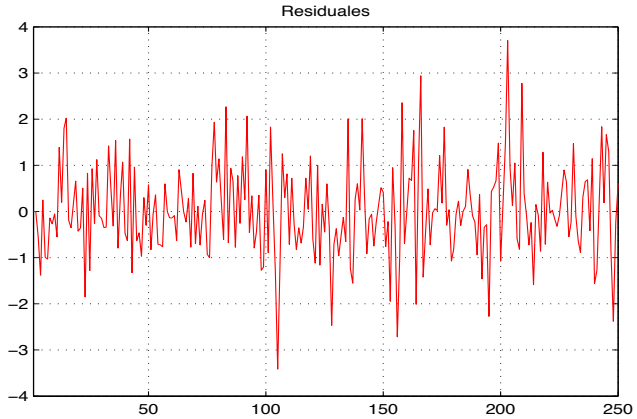
Ajuste de datos reales

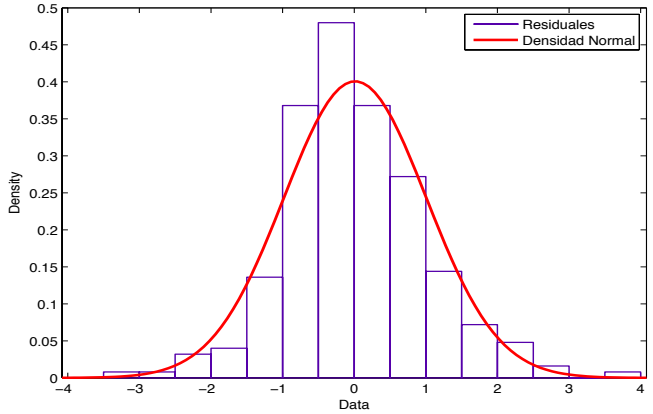
- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).
- Se ajustó la serie con el modelo propuesto y se comparó contra 3 modelos alternativos:
- Un AR(1) con errores distribuidos $N(0, \sigma^2)$.
- Un MA(1)-ARCH(1) con errores distribuidos $\mathcal{S}t(0, 1, \nu)$.
- AR(1)-ARCH(1) con errores distribuidos $\mathcal{S}t(0, 1, \nu)$.

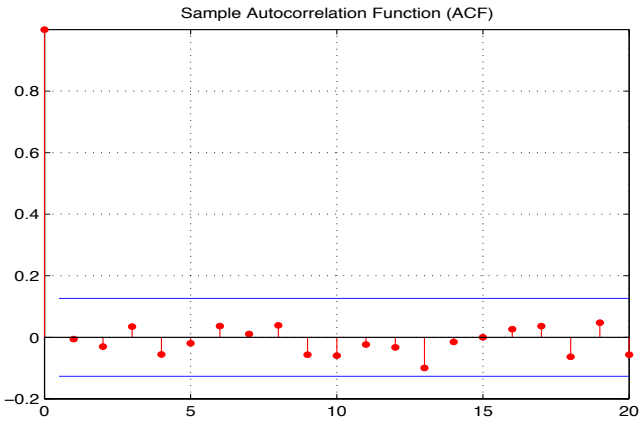
Ajuste de datos reales

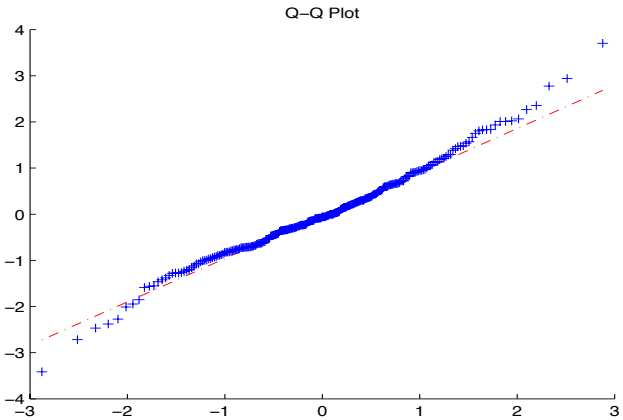
- Se modelaron los retornos históricos de la acción de AT&T Inc. Los datos son los precios diarios desde el 3 de octubre de 2005 hasta el 29 de septiembre de 2006 (251 datos).
- Se ajustó la serie con el modelo propuesto y se comparó contra 3 modelos alternativos:
- Un AR(1) con errores distribuidos $N(0, \sigma^2)$.
- Un MA(1)-ARCH(1) con errores distribuidos $\mathcal{S}t(0, 1, \nu)$.
- AR(1)-ARCH(1) con errores distribuidos $\mathcal{S}t(0, 1, \nu)$.
- Se realizaron 20 predicciones junto con sus intervalos calculados al 95 % de confianza y se compararon con las observaciones verdaderas.

Residuales del modelo AR(1) ordinario.

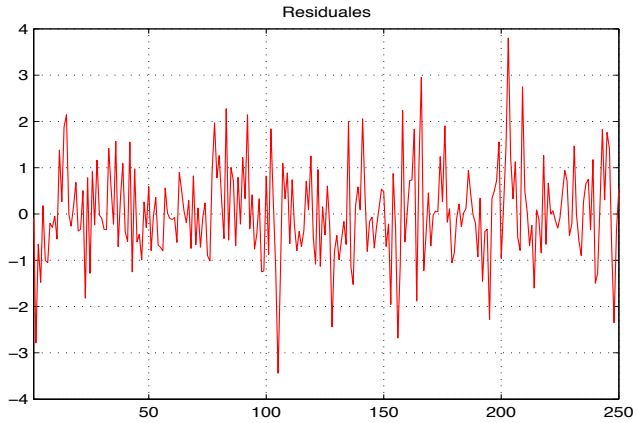


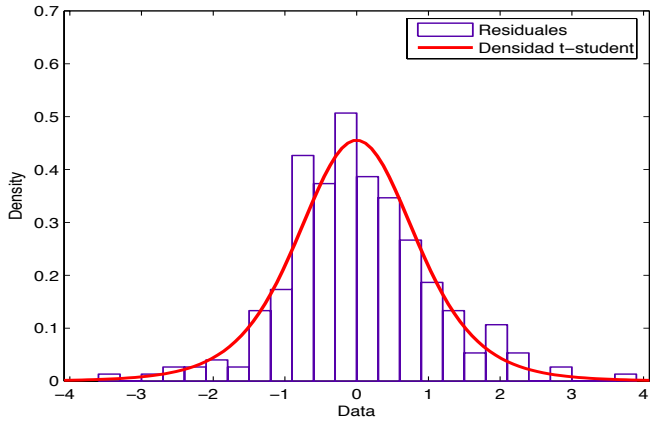


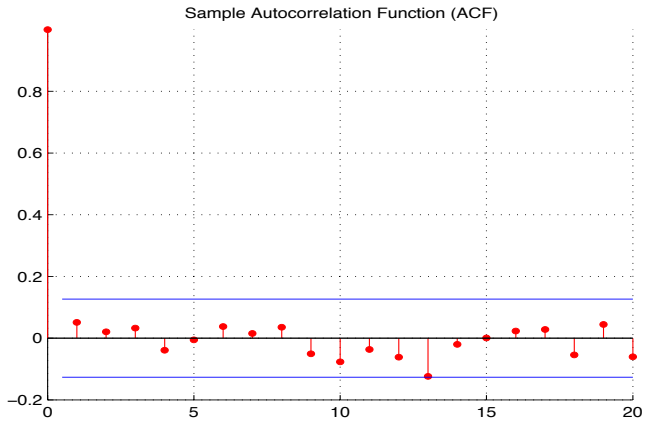


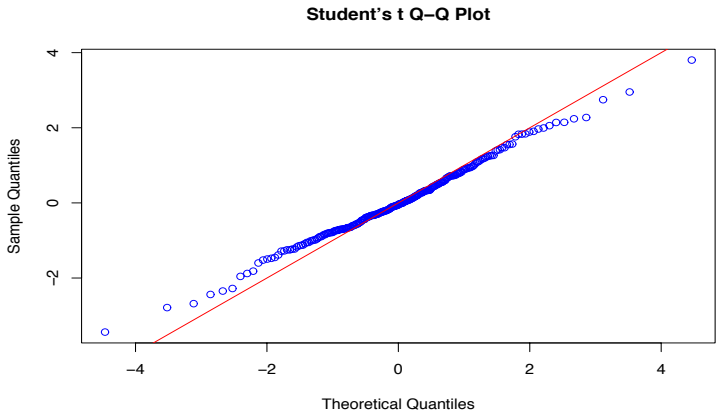


Residuales del modelo MA(1) - ARCH(1).

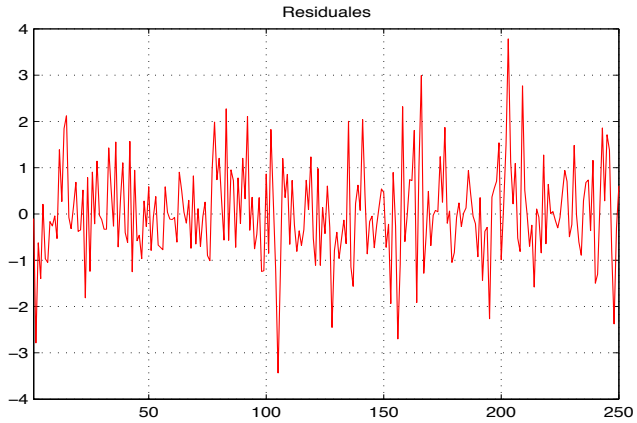


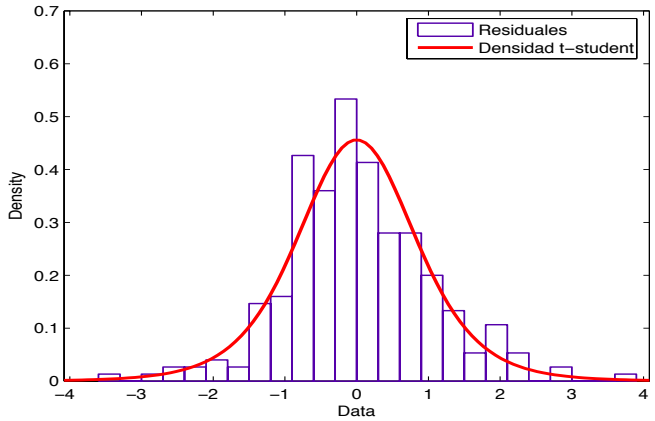


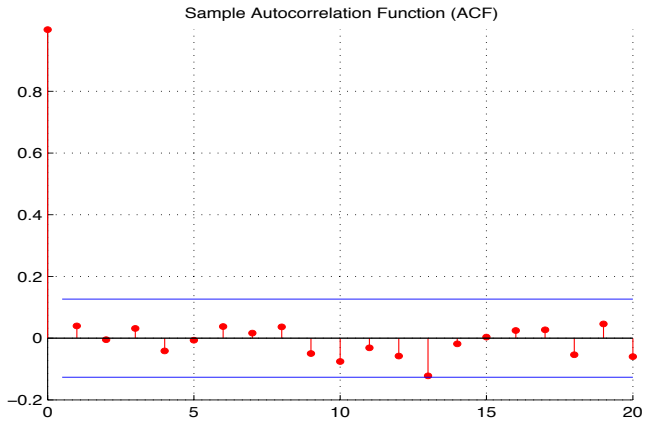


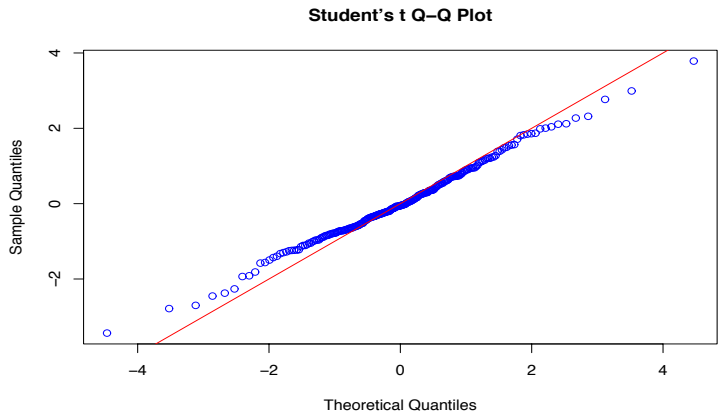


Residuales del modelo AR(1) - ARCH(1).

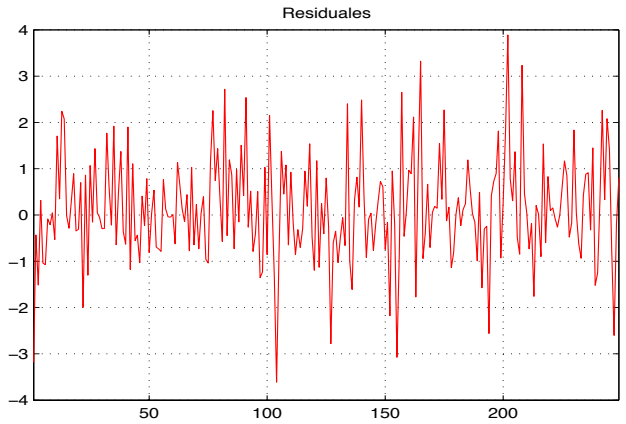


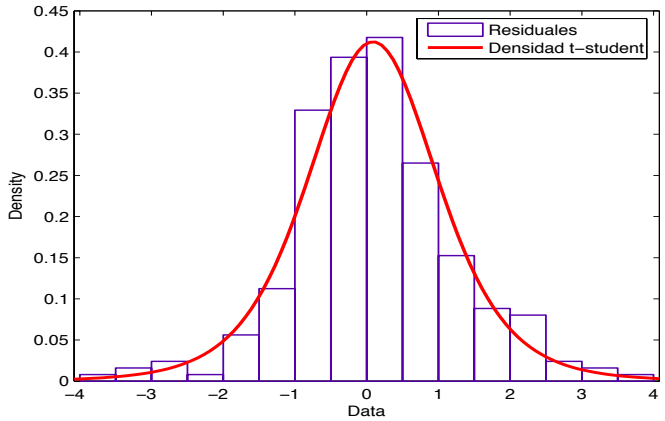


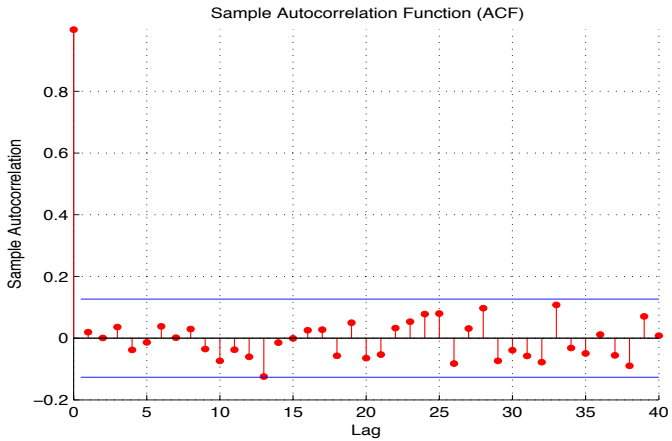


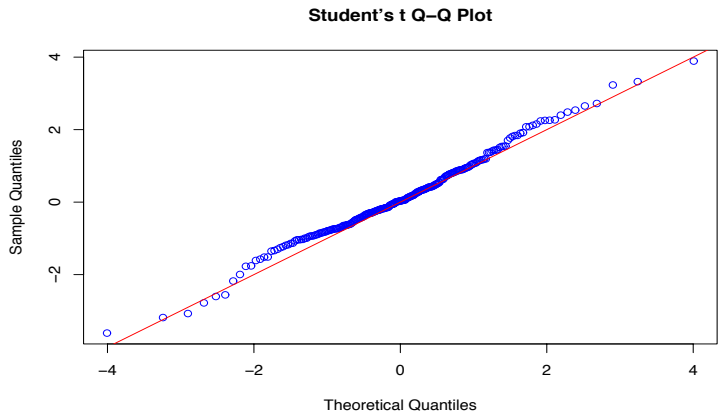


Residuales del modelo propuesto.









Resultados comparativos.

- Los modelos MA(1)-ARCH(1), AR(1)-ARCH(1) y AR(1)-ARCH(1) (usando variables latentes) pasaron las pruebas de bondad de ajuste y de no-correlación, sin embargo en el modelo AR(1) se rechazó la prueba de *Jarque-Bera* para probar Normalidad con lo cual queda descartado.

Resultados comparativos.

- Los modelos MA(1)-ARCH(1), AR(1)-ARCH(1) y AR(1)-ARCH(1) (usando variables latentes) pasaron las pruebas de bondad de ajuste y de no-correlación, sin embargo en el modelo AR(1) se rechazó la prueba de *Jarque-Bera* para probar Normalidad con lo cual queda descartado.
- Los otros MA(1)-ARCH(1), AR(1)-ARCH(1) que no consideran variables latentes en su construcción, presentaron deficiencias como exceso de curtosis respecto a la distribución t-Student, en sus histogramas vemos que los residuales no se asemejan mucho a una distribución t-Student, etc.

Resultados comparativos.

- Nuestro modelo tiene errores con distribución marginal $\mathcal{S}t(0, \beta^2, \nu)$ mientras que los modelos del tipo MA-ARCH y AR-ARCH contra los que se comparó, consideran una distribución $\mathcal{S}t(0, 1, \nu)$

Resultados comparativos.

- Nuestro modelo tiene errores con distribución marginal $\mathcal{S}t(0, \beta^2, \nu)$ mientras que los modelos del tipo MA-ARCH y AR-ARCH contra los que se comparó, consideran una distribución $\mathcal{S}t(0, 1, \nu)$
- 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.

Predicción para observaciones futuras.

- Para las predicciones se ajustaron 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.

Predicción para observaciones futuras.

- Para las predicciones se ajustaron 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.
- Se utilizó la ventaja de que los retornos $\{X_t\}_t$ sigue un proceso AR(1) para el cual la predicción para una observación futura x_{n+1} puede obtenerse mediante la ecuación

$$\hat{x}_{n+1} = \hat{\phi} x_n. \tag{6}$$

Predicción para observaciones futuras.

- El intervalo de confianza es análogo al intervalo de confianza para una predicción hecha mediante un modelo AR(1)

$$\hat{x}_{inf} = x_{pred} - \hat{\beta} \times \xi_{t_\nu}^{1-\alpha/2}, \ \hat{x}_{sup} = x_{pred} + \hat{\beta} \times \xi_{t_\nu}^{1-\alpha/2}, \quad (7)$$

Predicción para observaciones futuras.

- El intervalo de confianza es análogo al intervalo de confianza para una predicción hecha mediante un modelo AR(1)

$$\hat{x}_{inf} = x_{pred} - \hat{\beta} \times \xi_{t_\nu}^{1-\alpha/2}, \ \hat{x}_{sup} = x_{pred} + \hat{\beta} \times \xi_{t_\nu}^{1-\alpha/2}, \quad (7)$$

- De los 20 intervalos de confianza para las predicciones, 19 contuvieron al verdadero valor, es decir el 95 % como se deseaba.

Predicción para observaciones futuras.

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

Valuación de opciones con volatilidad estocástica

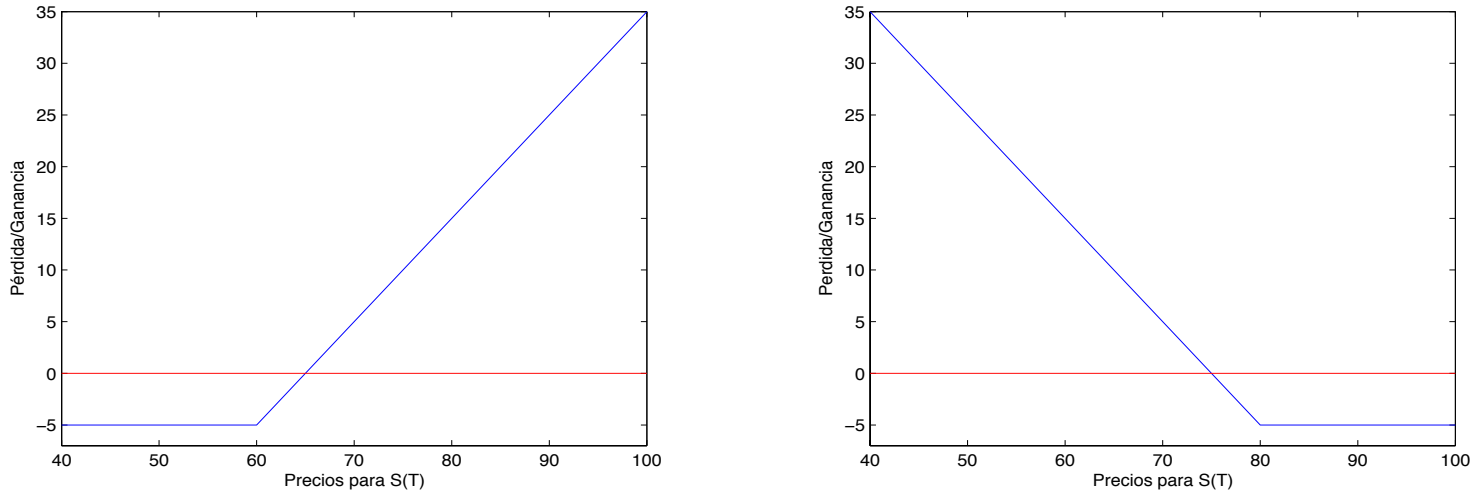


Figura: Diagramas de pérdidas y ganancias para una opción *call* (izquierda) y *put* (derecha).

Valuación de opciones con volatilidad estocástica

- 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

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

Valuación de opciones con volatilidad estocástica

- 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 $t = n+1, \ldots, \tau$

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

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

Valuación de opciones con volatilidad estocástica

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

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

Valuación de opciones con volatilidad estocástica

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

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

- El valor de la prima c para una opción *call* 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 proceso anterior.

Valuación de opciones con volatilidad estocástica

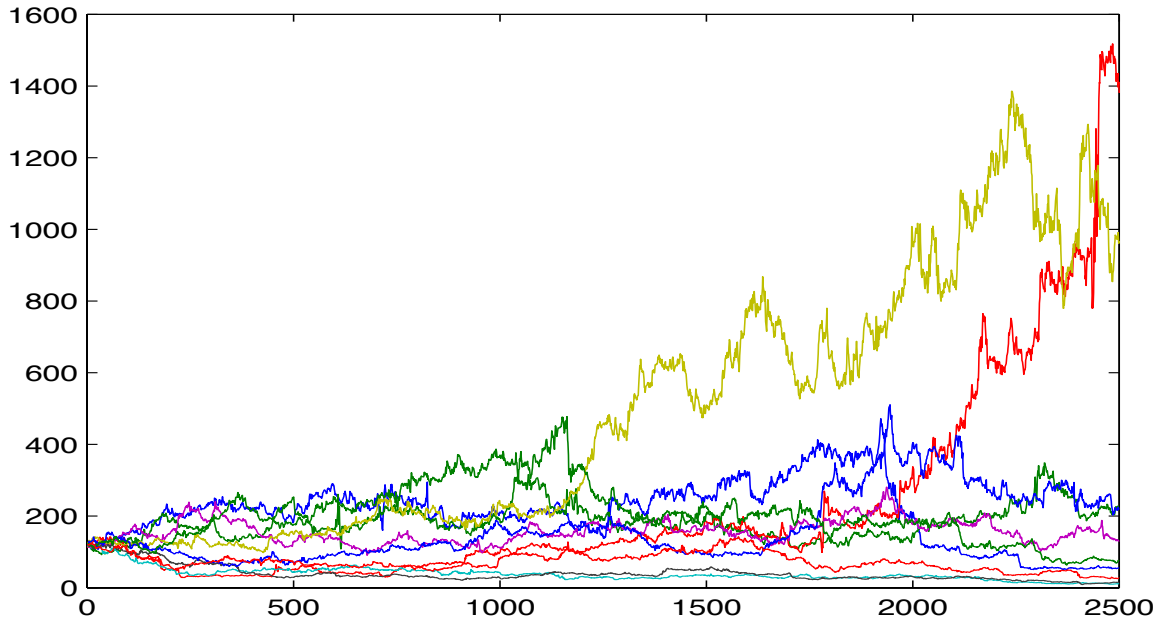


Figura: Simulación de 10 trayectorias de precios con retornos AR-ARCH.

Resultados comparativos

- Comparando el precio de una opción tipo *call* para la acción AT&T con respecto a otros modelos como son *Black-Scholes* y *Árboles Binomiales* se obtuvo lo siguiente:

$$PrecioCall_{BS} = 2.9615$$

$$PrecioCall_{AB} = 2.9798$$

$$PrecioCall_{Propuesto} = 2.9771$$

Resultados comparativos

- Comparando el precio de una opción tipo *call* para la acción AT&T con respecto a otros modelos como son *Black-Scholes* y *Árboles Binomiales* se obtuvo lo siguiente:

$$PrecioCall_{BS} = 2.9615$$

$$PrecioCall_{AB} = 2.9798$$

$$PrecioCall_{Propuesto} = 2.9771$$

- Como vemos los 3 precios son consistentes.

Conclusiones

- A lo largo de este trabajo hemos aprendido 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.

Conclusiones

- A lo largo de este trabajo hemos aprendido 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.
- Además se ilustró 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.

Conclusiones

- Como consecuencia de usar variables latentes para crear dependencia, nuestro modelo tiene errores con distribución marginal $\mathcal{S}t(0, \beta^2, \nu)$. Los modelos ARMA-GARCH tradicionales consideran una distribución $\mathcal{S}t(0, 1, \nu)$, la cual resulta menos flexible y no alcanza a capturar características importantes presentes en los datos como son la asimetría, que es una característica representativa de los modelos ARCH y GARCH. Consideremos que esta es una ventaja de usar nuestro esquema.

Conclusiones

- Ahora bien, pensemos que definiendo las distribuciones de transición apropiadas entre las observaciones y las variables latentes podemos especificar cualquier distribución de probabilidad, continua o discreta, como distribución marginal de nuestras observaciones.

Conclusiones

- Ahora bien, pensemos que definiendo las distribuciones de transición apropiadas entre las observaciones y las variables latentes podemos especificar cualquier distribución de probabilidad, continua o discreta, como distribución marginal de nuestras observaciones.
- A su vez, 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 estacionaridad de las observaciones. De esta manera, se podrían construir modelos de series de tiempo cada vez más complejos.

Bibliografía


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).


Bernardo, J. M. y Smith, A. F. M. (1994), *Bayesian Theory*. Chichester, U.K. Wiley.


Bernardo, J. M. (1980), *Bioestadística: Una Perspectiva Bayesiana*. Valencia, España. Vicens-Vives.


Bollerslev, T. (1986), *A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*. Journal of Econometrics 31, 307-327.


Bollerslev, T. (1987), *Generalized Autoregressive Conditional Heteroskedasticity*. The Review of Economics and Statistics, MIT Press, vol. 69(3), pages 542-47.


Brockwell, P. J. y Davis, R. A. (1987), *Time Series: Theory and Methods*. 2da Ed. Colorado, USA. Springer.


Brockwell, P. J. y Davis, R. A. (2002), *Introduction to Time Series and Forecasting*. 2da Ed. Colorado, USA. Springer.


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.


Contreras - Cristán, A.(2008), *Notas de Licenciatura del Curso: Series de Tiempo*. Facultad de Ciencias, UNAM.

Bibliografía


Dempster, A.P., Laird, N. y Rubin, D.B. (2003), *Maximun likelihood from incomplete data via the EM algorithm*. J. R. Statist. Soc. B, 39, 1-38.


Engle, R. (1982), *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, 1982, vol. 50, issue 4, pages 987-1007.


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.


Fan, J. y Yao, Q. (2003), *Nonlinear Time Series. Nonparametric and Parametric Methods*. Nueva York. Springer.


Gamerman, D., (1997), *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*. Londres. Chapman & Hall.


Gelman, A., Carlin, J., Stern, H. y Rubin, D. (1995), *Bayesian Data Analysis*. Londres. Chapman & Hall.


Glosten, L. R., Jagannathan, R. y Runkle, D.E. (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.


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.


Hull, J.C. y White, A. (1987), *The Pricing of Options on Assets With Stochastic Volatilities*. 42, 281-300. Journal of Finance.

Bibliografía


Hull, J.C.(2005), *Options, Futures and Other Derivatives*. 6ta Ed. Canadá. Prentice-Hall.


Kannan, D.(1979), *An Introduction to Stochastic Processes*. North Holland series in probability and applied mathematics.


Martinez, W. L. y Martinez, A. R (2002), *Computational Statistics Handbook with MATLAB*. USA. Chapman & Hall/CRC.


Nelson, D. B. (1991), *Conditional heteroskedasticity in asset returns: A new approach*. Econometrica 59: 347-370.


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.


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.


Robert, C. y Casella, G. (1999), *Monte Carlo Statistical Methods*. NY, USA. Springer.


Shumway, R. H. y Stoffer, D. S. (2006), *Time Series Analysis and Its Applications With R Examples*. 2da Ed. USA. Springer.


Stentoft L. (2004), *Pricing American Options when the Underlying Asset follows GARCH processes*. Dinamarca. School of Economics & Managment, University of Aarhus.

Bibliografía

Tanner, M.A.(1996), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3rd ed. Springer, New York.

Tsay, R. S. (2002), *Analysis of Financial Time Series. Financial Econometrics*. Candá. Wiley & Sons.

Wilmott, P. (2006), *Paul Wilmott on Quantitative Finance*. 2da Ed. Gran Bretaña. Wiley & Sons.

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.

Gracias por su atención.