MÓDULO DE ESTADÍSTICA SERIES DE TIEMPO

DIPLOMADO DE MATEMÁTICAS Y ESTADÍSTICA, MIDE

Ruslán Gómez Nesterkín

Noviembre 2022

Aviso: Los comentarios y opiniones expresados son solo del autor y no necesariamente reflejan a los del Banco de México.

ENCUESTAS

- A fin de tener una mayor interacción durante las presentaciones…
- Se realizarán algunas encuestas de opinión durante las sesiones…
- A través de la aplicación de celular: **Telegram**
- Favor de adherirse al grupo: **MIDE_DIP_MATS_2022**
- Liga directa: *https://t.me/MIDE_DIP_MATS_2022*

PLAN DE LA PRESENTACIÓN

PARTE 1

- **ANTECEDENTES:** Series determinísticas, variables aleatorias y estimación.
- **SERIES DE TIEMPO:** Series Determinísticas, Series de Tiempo, Suma de Variables Aleatorias.
- **DISTRIBUCIÓN DE SERIES DE TIEMPO:** Límite Central, Función Característica, Caminata Aleatoria.
- **CARACTERÍSTICAS DE LAS SERIES DE TIEMPO:** Autocorrelación, Estacionalidad.

PARTE 2

- **DESCOMPOSICIÓN:** Descomposición de Wold.
- **MODELOS DE SERIES DE TIEMPO:** Función de auto-correlación, AR, MA.
- **EJEMPLO:** ¿Predicción al 75%?

PARTE 3

- **MODELOS DE SERIES DE TIEMPO:** Función de auto-correlación parcial, ARMA, otras variantes.
- **OTRAS CARACTERÍSTICAS:** Heteroskedasticidad, Modelo ARCH, Causalidad de Granger, Cointegración.
- **PRONÓSTICOS:** ARMA / Box-Jenkins.

PARTE 3
PRONÓSTICOS

PLAN DE LA PRESENTACIÓN

PARTE 1

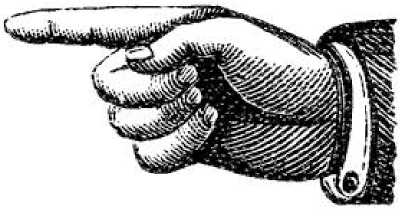
- ANTECEDENTES
- SERIES DE TIEMPO
- DISTRIBUCIÓN DE SERIES DE TIEMPO
- CARACTERÍSTICAS DE LAS SERIES DE TIEMPO

PARTE 2

- DESCOMPOSICIÓN
- MODELOS DE SERIES DE TIEMPO
- EJEMPLO

PARTE 3

- MODELOS DE SERIES DE TIEMPO
- OTRAS CARACTERÍSTICAS
- PRONÓSTICOS

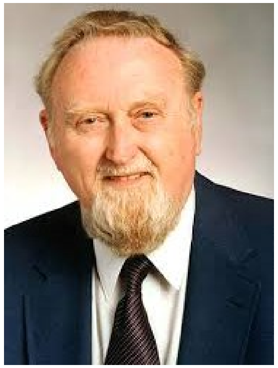


MODELOS

Motivación para análisis de series de tiempo

MODELOS ESTOCÁSTICOS

Robert F. Engle III
Clive W. J. Granger



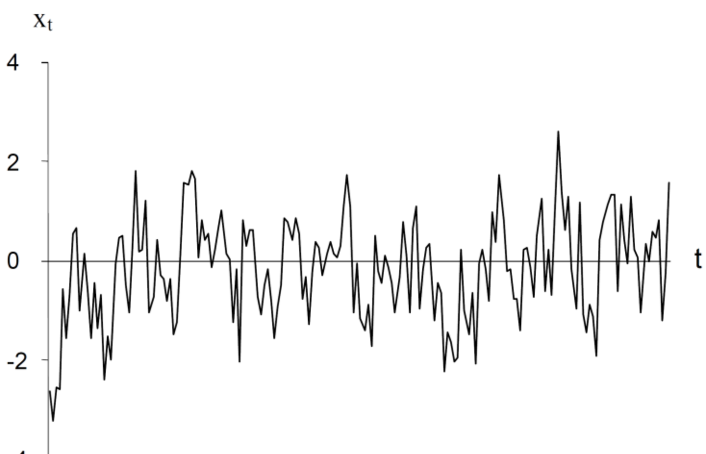
Premio Nobel 2003 en Ciencias Económicas *"por métodos para el análisis económico de series de tiempo con volatilidad variable en el tiempo (ARCH)."*

https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2003/press.html

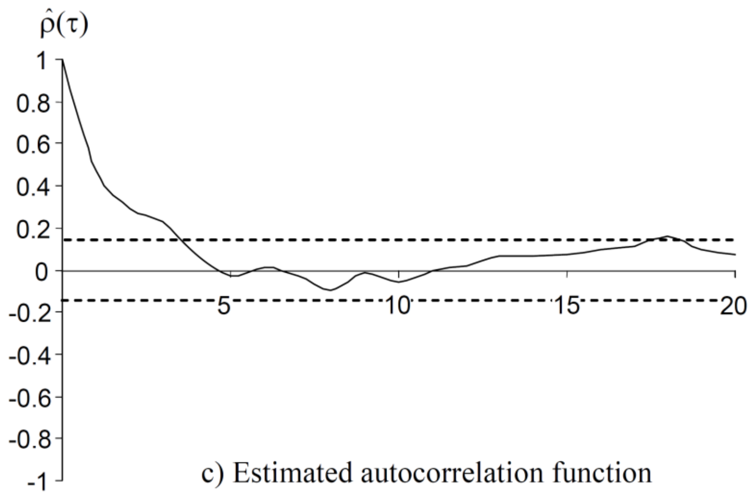
- Métodos (matemáticos) para análisis de series de tiempo.
- Modelos (matemáticos) de volatilidad estocástica.
- Modelos (estadísticos) con memoria de largo plazo.
- Metodologías (estadísticas) para identificar causalidad.
- Muchas de las series de tiempo económicas son no-estacionarias.

Función de Autocorrelación Parcial (PACF) (1/2)

- La modelación con AR(p) conduce a la pregunta de cuántos términos p utilizar.
- Una característica de los modelos AR(p) en procesos estacionarios, es que su función de auto-correlación ρ_τ, como función del rezago τ, converge asintóticamente a cero.
- Ejemplo: Modelo AR(1) de un proceso aleatorio puro de 240 observaciones con $\alpha = 0.5$:



a) Realisation



c) Estimated autocorrelation function with confidence intervals

Función de Autocorrelación Parcial (PACF) (2/2)

- Asumamos una serie de valores observados $x = \{x_1, \ldots, x_n\}$ y los modelos AR(p) siguientes:

$$
\begin{aligned}
X_t &= \phi_{0,1} + \boxed{\phi_{1,1}} X_{t-1} + e_{1,t} && p = 1 \\
X_t &= \phi_{0,2} + \phi_{1,2} X_{t-1} + \boxed{\phi_{2,2}} X_{t-2} + e_{2,t} && p = 2 \\
X_t &= \phi_{0,3} + \phi_{1,3} X_{t-1} + \phi_{2,3} X_{t-2} + \boxed{\phi_{3,3}} X_{t-3} + e_{3,t} && p = 3 \\
&\vdots
\end{aligned}
$$

- Para cada AR(p) es posible estimar los parámetros $\phi_{i,j}$ usando los datos x, por ejemplo mediante mínimos cuadrados, obteniendo los estimadores $\hat{\phi}_{i,j}$.

- El estimador $\hat{\phi}_{i,i}$ representa el peso del término X_{t-i}, de modo que como la auto-correlación es asintótica a cero, se esperaría que a partir de un cierto p, el estimador $\hat{\phi}_{p+s,p+s}$ sea muy cercano a cero.

- **Criterio PACF:** Dada una tolerancia $\varepsilon > 0$ (nivel de tolerancia preconcebido), escogemos el modelo AR(p) como aquel a partir del cual el coeficiente $\hat{\phi}_{i,i}$ sea despreciable:

$$p^* = \min\{p : |\hat{\phi}_{p+s,p+s}| < \varepsilon\}$$

Nota: Este criterio evita sobrecarga de variables (**overfitting**) usadas en el modelo.

Promedio Móvil Auto Regresivo:
(ARMA)

- Un ARMA(p, q) se refiere a una combinación de un AR(p) y un MA(q).

$X_t = \quad \delta$ ← Término relativo a **tendencia**

$+\alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p}$ ← Término relativo a **AR(p)**

$+u_t$ ← Término relativo a **aleatoriedad** (en t)

$-\beta_1 u_{t-1} - \cdots - \beta_q u_{t-q}$ ← Término relativo a **MA(q)**

- De esta manera obtenemos el modelo **ARMA(1,1)**:

$$\boxed{X_t = \delta + \alpha X_{t-1} + u_t - \beta u_{t-1}}$$

EJERCICIO EN CLASE (10 minutos): En un modelo **ARMA(1,1)**, sean $\alpha = -0.8, \beta = 0.5, u_t \sim \boldsymbol{N}(0,1)$ y $\delta = 0$. Si se observan los valores de la realización de u_t siguientes: $u_0 = 1, u_1 = -1$ y $u_2 = 0.5$, ¿cuál sería el valor que tendrían X_1 y X_2 cuando $X_0 = 0$?

Responder en el grupo MIDE_DIP_MATS de TELEGRAM. [MIDE_ST_P3_a]

Promedio Móvil Auto Regresivo: Momentos de ARMA(1,1) (1/2)

- Calculamos el **valor esperado** del proceso estacionario X_t:

1. Por definición: $\mathbb{E}[X_t] = \mathbb{E}[\delta + \alpha X_{t-1} + u_t - \beta u_{t-1}] = \delta + \alpha \mathbb{E}[X_{t-1}]$

2. Como X_t es estacionario, entonces tiene el mismo valor esperado para todo t:

$$\mu = \mathbb{E}[X_t] = \mathbb{E}[X_{t-1}]$$

3. Tomando los dos resultados anteriores, resolvemos μ de la siguiente ecuación:

$$\mu = \delta + \alpha\mu \quad \Longrightarrow \quad \mu = \frac{\delta}{(1-\alpha)}$$

- Por lo tanto el valor esperado de un ARMA(1,1) es el mismo de un AR(1).

$$\mathbb{E}[X_t] = \frac{\delta}{(1-\alpha)}$$

Promedio Móvil Auto Regresivo: Momentos de ARMA(1,1) (2/2)

- Calculamos la auto-correlación del proceso estacionario X_t:

$$\rho_\tau = \mathbb{E}[X_{t-\tau}\, X_t] = \mathbb{E}[X_{t-\tau}(\delta + \alpha X_{t-1} + u_t - \beta u_{t-1})]$$

- Obteniendo la **función de autocorrelación** ρ_τ como una fórmula recursiva:

$$\begin{cases} \rho_\tau = \alpha \rho_{\tau-1} \\ \rho_1 = \dfrac{(\alpha - \beta)(1 - \alpha\beta)}{1 + \beta^2 - 2\alpha\beta} \end{cases}$$

- **PREGUNTAS:**
  - ¿Qué sucede con ρ_τ cuando $\alpha = \beta$?
  - ¿Qué podemos decir entonces de un modelo ARMA(1,1) X_t cuando $\alpha = \beta$?

EJERCICIO EN CLASE (5 minutos): Obtener la varianza (equivalente al coeficiente de autocorrelación ρ_0) de un **ARMA(1,1)** cuando $\alpha = 1$ y $\beta = 0$.

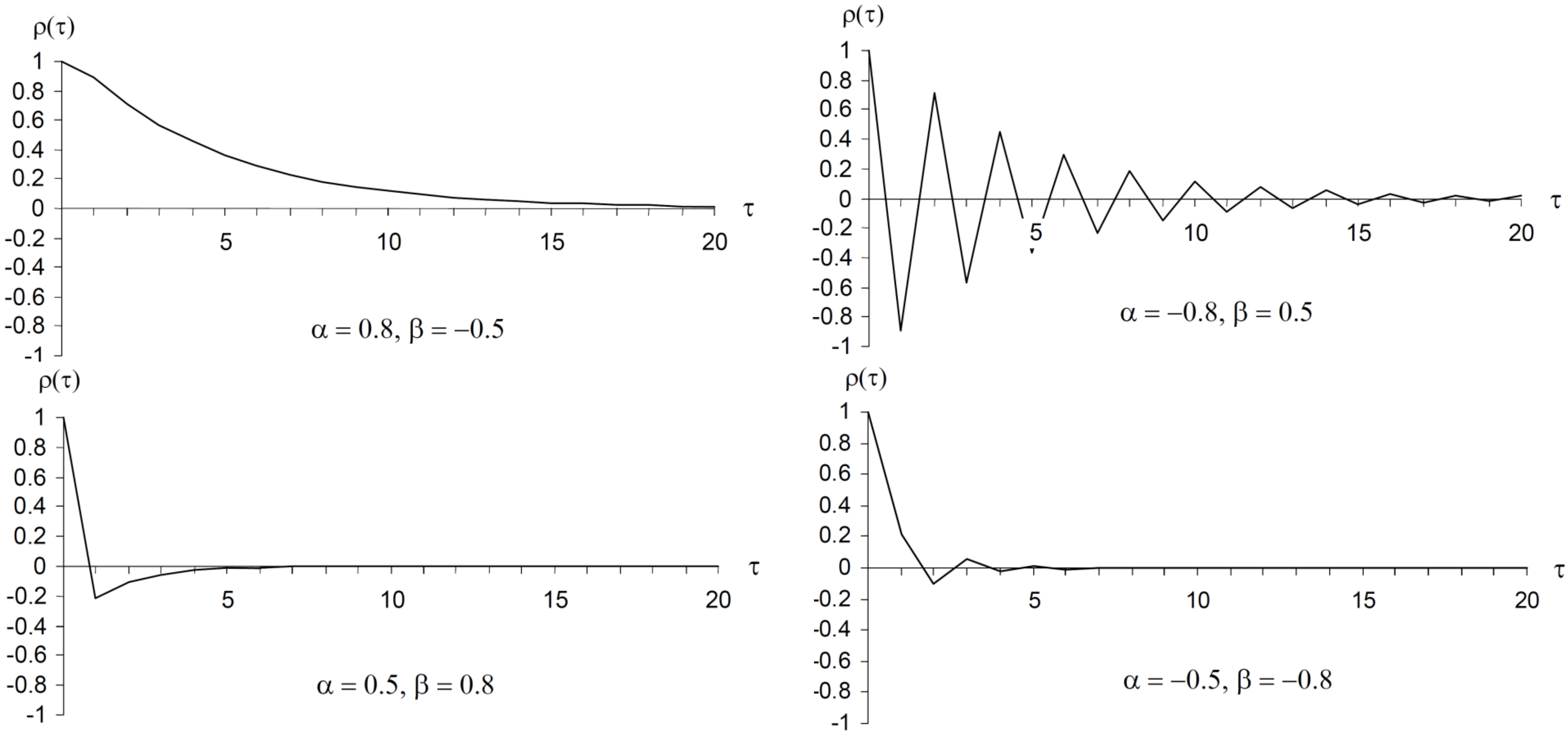
Responder en el grupo MIDE_DIP_MATS de TELEGRAM. [MIDE_ST_P3_b]

MODELOS

(Modelo: $X_t = \delta + \alpha X_{t-1} + u_t - \beta u_{t-1}$)

Promedio Móvil Auto Regresivo (2/2)
Momentos de ARMA(1,1)

Ejemplo: Considérese a un proceso ARMA(1,1) generado con u_t v.a.'s i.i.d dadas por procesos aleatorios puros. Se muestran a continuación las funciones de auto-correlación ρ_τ para diferentes combinaciones de α y β tomando valores 0.8, 0.5, -0.5 y -0.8.



5 minutos ...

PLAN DE LA PRESENTACIÓN

PARTE 1

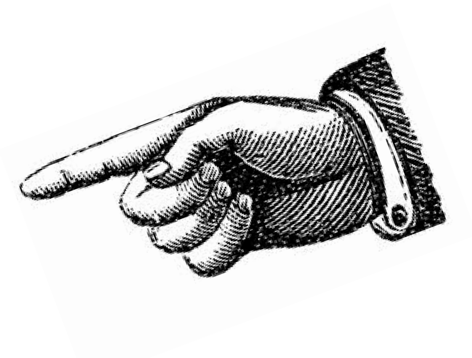
- ANTECEDENTES
- SERIES DE TIEMPO
- DISTRIBUCIÓN DE SERIES DE TIEMPO
- CARACTERÍSTICAS DE LAS SERIES DE TIEMPO

PARTE 2

- DESCOMPOSICIÓN
- MODELOS DE SERIES DE TIEMPO
- EJEMPLO

PARTE 3

- MODELOS DE SERIES DE TIEMPO
- OTRAS CARACTERÍSTICAS
- PRONÓSTICOS



OTRAS CARACTERÍSTICAS

Causalidad (de Granger)

- **Correlación:** Al considerar a dos v.a.'s X y Y, su correlación refleja que cambios de una v.a. se deben reflejar también en la otra.
- **Correlaciones espurias:** Hay ejemplos para los cuales la correlación no hace sentido y se denominan espurias.
- **Alternativa:** Concepto de causalidad.
- **Causalidad a la Granger**:
  - Los modelos ARMA para X_t, tienen como consideración implícita el uso óptimo de información pasada $\mathrm{D}_t^x = \{x_{t-1}, x_{t-2}, \dots, x_0\}$.
  - La idea consiste en que si un proceso X causa a Y, los valores actuales y rezagados de X_t deben contener información D_t^x que al usarse permita mejorar la predicción de $Y_{t+\tau}$, para $\tau > 0$.
  - Otra forma de entender la idea anterior es que la información para predecir $Y_{t+\tau}$ no está solo contenida en su historia, sino también en la historia o distribución D_t^x del proceso X_t.

OTRAS CARACTERÍSTICAS

Cointegración: Grado de Integración

- Partamos de la regresión lineal siguiente, con u_t proceso aleatorio puro:

$$y_t = \alpha + \beta x_t + u_t$$

Variable observada

Modelo lineal usado

- Tenemos que y_t depende de x_t y del ruido generado por u_t.
- Conforme a Granger (1981), una simulación del lado derecho debe reproducir las propiedades de la variable observada y_t.
- Si por ejemplo y_t muestra ciclicidad, entonces x_t debe presentar también dicha ciclicidad, ya que la aleatoriedad de u_t no la incorporaría.
- Se define el concepto de **grado de integración** $\mathrm{I}(d)$ a partir de la idea de que diferenciando d veces una variable y_t pueda hacerse estacionaria. En tal caso se dirá que y_t tiene **grado de integración de orden** d:

Si $y_t \sim \mathrm{I}(d)$ significa que $\Delta_d y_t$ es estacionaria.

OTRAS CARACTERÍSTICAS

Cointegración: Método de Representación de Granger

- **Propiedades** del Grado de Integración :
  - Si y_t es débilmente estacionaria entonces $y_t \sim \mathrm{I}(0)$.
  - Si $y_t \sim \mathrm{I}(1)$ entonces $\Delta_1 y_t = y_t - y_{t-1} \sim \mathrm{I}(0)$.
  - Si $y_t \sim \mathrm{I}(1)$ y $z_t \sim \mathrm{I}(0)$ entonces $y_t + z_t \sim \mathrm{I}(1)$ normalmente.

- **Variables Cointegradas:** Se refiere a que dos procesos $\mathrm{I}(1)$ al sumarlos resulten $\mathrm{I}(0)$.

- **Ejemplo:** Sea $y_t = \alpha + \beta x_t + u_t$ con u_t proceso aleatorio puro y supongamos que $y_t \sim \mathrm{I}(1)$ y $x_t \sim \mathrm{I}(1)$. Entonces:

  1. Sabemos que $u_t \sim \mathrm{I}(0)$ por ser estacionario.
  2. Resulta que $y_t - \beta x_t \sim \mathrm{I}(0)$ por como definimos a y_t.
  3. Del caso anterior, β es única para que se presente esta situación, ya que:

  Asumiendo que $y_t = \beta_1 x_t + u_{1,t}$ y $y_t = \beta_2 x_t + u_{2,t}$, entonces resulta que $\beta_2 = \beta_1$:

$$(\beta_2 - \beta_1) x_t = u_{1,t} - u_{2,t}$$

$\mathrm{I}(1)$ a menos que $\beta_2 = \beta_1$

$\mathrm{I}(0)$ ya que la diferencia es estacionaria

OTRAS CARACTERÍSTICAS

Teorema de Representación de Granger

- **Definición formal de causalidad de Granger (1969):**

  - **Supuestos:** Sea $\mathrm{D}_t^x = \{x_t, x_{t-1}, \dots, x_0\}$ la información de proceso X_t al tiempo t, D_t^y la correspondiente al proceso Y_t y $\mathrm{D}_t^* = \{x_t, \dots, x_0, y_t, \dots, y_0\}$ la información total de ambos procesos.

  - **Métrica para errores:** Sea $\sigma^2(y_{t+1})$ la varianza del error por predicción de y_{t+1} en el tiempo t:
  $$\sigma^2(y_{t+1}) = \mathbb{E}[(\hat{y}_{t+1} - y_{t+1})^2]$$

  - **Causalidad de Granger**: x causa a y si y solo si la predicción óptima lineal cumple:
  $$\sigma^2(y_{t+1}|\mathrm{D}_t^*) < \sigma^2\left(y_{t+1}|\mathrm{D}_t^y\right)$$

- **Interpretación:** Los valores de y pueden predecirse mejor al adicionar la información actual y anterior de x.

OTRAS CARACTERÍSTICAS

Teorema de Representación de Granger

Ilustración del Teorema de Representación de Granger y Weiss (1983)

- Considérese el siguiente sistema de **AR**(p), con $u_{1,t}$ y $u_{2,t}$ procesos aleatorios puros:

$$x_t = \sum_{j=1}^{p} a_{1,j} x_{t-j} + \sum_{j=1}^{p} b_{1,j} y_{t-j} + u_{1,t}$$

$$y_t = \sum_{j=1}^{p} a_{2,j} x_{t-j} + \sum_{j=1}^{p} b_{2,j} y_{t-j} + u_{2,t}$$

- Si x_t y y_t son $\mathrm{I}(1)$ y **cointegradas**, entonces el sistema puede reescribirse como:

$$\Delta x_t = \alpha_1 (y_{t-1} - \beta x_{t-1}) \sum_{j=1}^{p-1} a_{1,j}^{*} \Delta x_{t-j} + \sum_{j=1}^{p-1} b_{1,j}^{*} \Delta y_{t-j} + u_{1,t}$$

$$\Delta y_t = \alpha_2 (y_{t-1} - \beta x_{t-1}) \sum_{j=1}^{p-1} a_{2,j}^{*} \Delta x_{t-j} + \sum_{j=1}^{p-1} b_{2,j}^{*} \Delta y_{t-j} + u_{2,t}$$

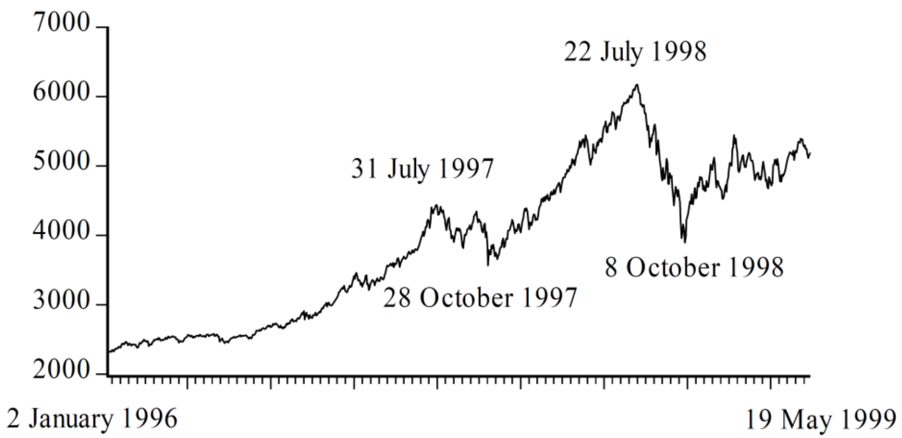
con al menos algún α_1 o α_2 distinto de cero.

- En este caso se tiene que ambas partes de las ecuaciones están equilibradas y tienen orden de integración $\mathrm{I}(0)$.

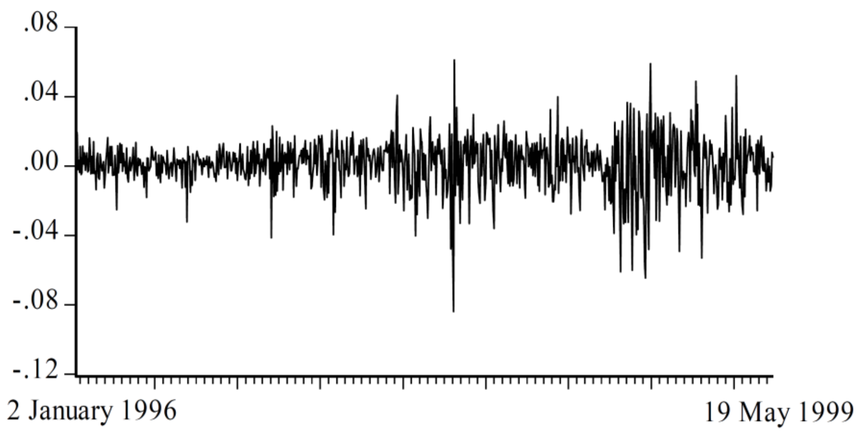
OTRAS CARACTERÍSTICAS

Heteroskedasticidad: Antecedentes

- **Heteroskedasticidad** proviene del griego antigüo ***hetero*** "diferente" y ***skedasis*** "dispersión" (*https://en.wikipedia.org/wiki/Heteroscedasticity*)
- La **heteroskedasticidad** es una característica observada en muchas series de tiempo de datos reales, con la volatilidad (dispersión) variando en el tiempo.
- **Ejemplo:** Tomemos el índice accionario del mercado alemán (DAX), consistente en 842 observaciones diarias del 2/01/1996 al 19/05/1999:



a) German Stock Market Index: Data



b) German Stock Market Index: Continuous returns

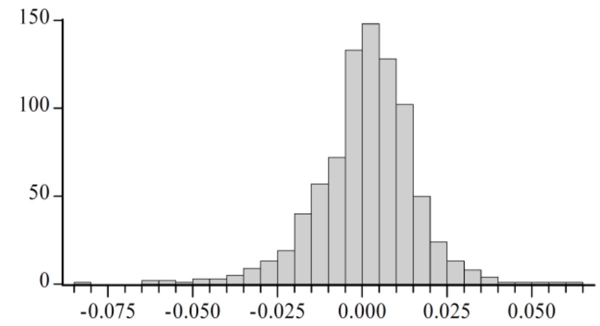
OTRAS CARACTERÍSTICAS

Heteroskedasticidad: Ejemplo

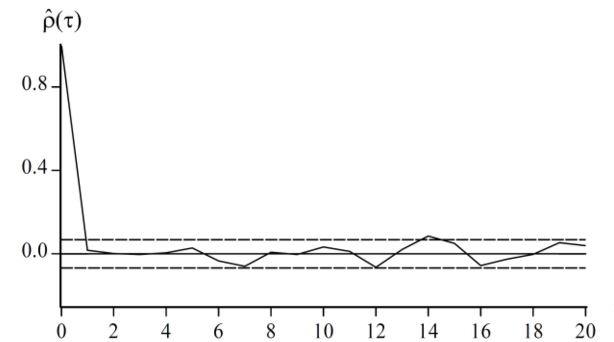
- Se aprecia en el histograma que la distribución de los rendimientos presenta kurtosis de 6.3, superior al 3.0 que se esperaría para una distribución normal.
- El correlograma o ACF muestra un **orden 2** de auto-correlación.
- Al estimar un modelo **AR(2)** para los rendimientos observados, se obtienen la siguiente estimación:

$$\Delta \ln(\mathrm{DAX}_t) = 0.001 - 0.090\, \Delta \ln(\mathrm{DAX}_{t-2}) + \varepsilon_t$$

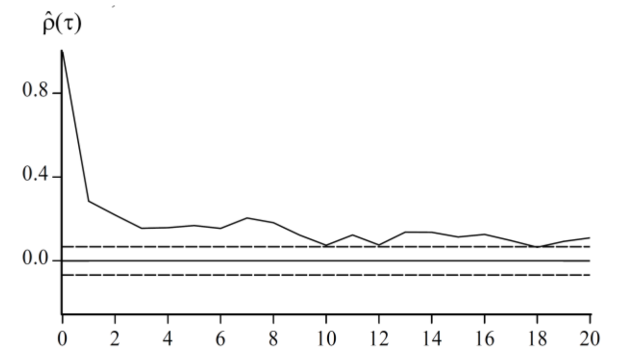
- Pero cuando se calcula la ACF para el cuadrado de los residuos ε_t^2, se observa que los datos no son estacionarios. Esto indica que hay un **segundo momento (varianza) con autocorrelación.**



c) German Stock Market Index: Histogram of the continuous returns



d) Estimated autocorrelations of the residuals



e) Estimated autocorrelations of the squared residuals

Modelo ARCH: Introducción
(Auto Regresive Conditional Heteroskedasticity)

- Un ejemplo de modelo ARCH estaría dado por la siguiente relación lineal:

$$y_t = \beta x_t + \varepsilon_t$$

- Con la posibilidad de que la variable x_t pueda contener variables endógenas con rezago:

$$x_t = \sum_{j=1}^{p} a_j \, x_{t-j}$$

- El término de error ε_t tiene media cero y varianza estacionarias:

$$\mathbb{E}[\varepsilon_t] = 0, \qquad \mathbb{E}[\varepsilon_t] = \sigma^2$$

- Se asume que ε_t no es auto-correlacionada, sin embargo ${\varepsilon_t}^2$ si presenta auto-correlación, representada por un modelo AR(q):

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_0 \varepsilon_{t-j}^2 + \omega_t$$

con ω_t un proceso aleatorio puro, $\alpha_0 > 0, \alpha_q > 0$ y $\alpha_i \geq 0$ para $i = 1, \dots q - 1$.

PLAN DE LA PRESENTACIÓN

PARTE 1

- ANTECEDENTES
- SERIES DE TIEMPO
- DISTRIBUCIÓN DE SERIES DE TIEMPO
- CARACTERÍSTICAS DE LAS SERIES DE TIEMPO

PARTE 2

- DESCOMPOSICIÓN
- MODELOS DE SERIES DE TIEMPO
- EJEMPLO

PARTE 3

- MODELOS DE SERIES DE TIEMPO
- OTRAS CARACTERÍSTICAS
- PRONÓSTICOS

Predicción con ARMA: mediante mínimos cuadrados. (1/2)

- En los años 1970 tuvo gran auge el enfoque predictivo propuesto por **Box-Jeknins** que se describe a continuación debido a la simplicidad del método comparado otros modelos econométricos con muchas variables, resultando estos últimos más engorrosos.

- Se asumirá que todas las observaciones previas y hasta el tiempo t son conocidas:

Serie de datos conocidos: $\{x_t, x_{t-1}, \dots, x_0\}$

- Denotaremos a $\hat{X}_t(\tau)$ como el estimador en el periodo t, de la serie τ periodos en el futuro correspondientes a $x_{t+\tau}$.

- El estimador $\hat{X}_t(\tau)$ se obtiene de forma recursiva...

Predicción con ARMA: mediante mínimos cuadrados. (2/2)

- Realizando la predicción a un periodo, quedaría el estimador de la siguiente manera:

$$\hat{X}_t(1) = \delta + \alpha x_t - \beta u_t$$

- Si consideramos que

$$\hat{X}_0(1) = \delta + \alpha x_0 - \beta u_0$$

Tomado al error de la predicción $u_1 = x_1 - \hat{X}_0(1)$, resulta al combinarlos que:

$$\hat{X}_1(1) = \delta(1+\beta) + (\alpha - \beta)x_1 + \alpha\beta x_0 - \beta^2 u_0$$

- Procediendo de esta forma para los siguientes periodos de tiempo t se obtiene que:

$$\boxed{\hat{X}_t(1) = \delta \sum_{j=0}^{t} \beta^j + \sum_{j=0}^{t-1} \beta^j (\alpha - \beta)\, x_{t-j} + \beta^t \alpha x_0 - \beta^{t+1} u_0}$$

CONCLUSIONES

REFLEXIONES FINALES

- **Series de tiempo e inferencia estadística:** Se volvieron un estándar en la modelación de fenómenos económicos o econometría (**Haavelmo et. al.**) y en otras disciplinas también (medicina, farmacéutica, psicología, etc.).
- **Modelos Autoregresivos:** Resultaron ser útiles para representar fenómenos económicos más complejos (**Engle** y **Granger et. al.**).
- **Predicción:** Se antoja extrapolar al futuro con esto modelos... Causalidad vs Predicción (**Granger et. al.**)
- **Paradojas sobre predicción:** Es fundamental la independencia de las v.a.'s (ejemplo de **Sornette y Andersen**). Si se pudiera predecir ahora, cualquier incorporación de dicha predicción en la práctica implicaría cambios en el futuro, afectándose la veracidad de la predicción original...
- **Conclusión:** ... ¡Hay que tomar en cuenta la información disponible! ... ¡y saber utilizarla!

"Si sale sol, le ponemos un bypas cuádruple. Si cae águila, se toma un desenfriolito."

https://www.physiciansweekly.com/aspirin-cartoon/

MÓDULO DE ESTADÍSTICA SERIES DE TIEMPO

DIPLOMADO DE MATEMÁTICAS Y ESTADÍSTICA, MIDE

Ruslán Gómez Nesterkín

Noviembre 2021