



Estos apuntes fueron preparados para el curso

EC4301 - Macroeconometría

Randall Romero Aguilar, PhD
randall.romero@ucr.ac.cr

II Semestre 2020

Preliminar

Estos apuntes fueron actualizados por última vez el 10 de agosto de 2020. Estaré agregando material nuevo a este texto durante 2020. Una versión más reciente podría estar disponible en <http://randall-romero.com/ec4301-2020-ii/>



Índice general

I	Métodos para el análisis de una serie de tiempo individual	1
1	Introducción al análisis de series de tiempo	3
2	Ecuaciones en diferencia	25
3	Modelos AutoRegresivos de Media Móvil (ARMA)	39
4	Modelos de volatilidad	69
5	Modelos de tendencia	87
6	Cambio estructural	113
7	Modelos de estacionalidad	127
II	Métodos para el análisis conjunto de varias series de tiempo	141
8	Introducción a la estimación de sistemas de ecuaciones	143
9	Sistemas de ecuaciones simultáneas	149
10	Vectores Auto-Regresivos (VAR)	167
11	Cointegración y vector de corrección de errores (VECM)	187

Todavía está incompleto

Referencias propuestas para estos apuntes:

- Walter Enders (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6
- Gebhard Kirchgässner, Jürgen Wolters y Uwe Hassler (2013). *Introduction to Modern Time Series Analysis*. 2^a ed. Springer. ISBN: 978-3-642-33435-1

- George E. P. Box y col. (2016). *Time Series Analysis*. 5^a ed. Wiley. ISBN: 978-1-118-67502-1
- Terence C. Mills (2019). *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*. Academic Press. ISBN: 978-0-12-813117-6
- Gunnar Bardsen y col. (2005). *The Econometrics of Macroeconomic Modelling*
- Helmut Lütkepohl y Markus Kratzig, eds. (2004). *Applied Time Series Econometrics*. Cambridge University Press. ISBN: 978-0-521-83919-8
- James M. Hamilton (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6

Copyright

This material is shared under a **Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)** Creative Commons license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



**Attribution-NonCommercial-ShareAlike
4.0 International (CC BY-NC-SA 4.0)**

Parte I

Métodos para el análisis de una serie de tiempo individual

Capítulo 1

Introducción al análisis de series de tiempo

En este capítulo

1.1	Una breve historia	3
1.2	Representación gráfica de series de tiempo	6
1.3	El operador de rezagos	11
1.4	Procesos estocásticos	13
1.5	Series de tiempo	15

1.1 Una breve historia

¿Para qué estudiar series de tiempo

El análisis de series de tiempo ayuda a detectar *regularidades* en las observaciones de una variable y derivar “leyes” a partir de ellas, o bien para explotar toda la información incluida en la variable para *predecir* mejor el futuro.

- El análisis de series de tiempo ha tenido un papel importante en la ciencia desde la antigüedad.
 - Por ejemplo, los astrónomos babilonios usaron series de tiempo de las posiciones relativas de las estrellas y planetas para predecir eventos astronómicos.
 - Kepler descubrió las leyes que llevan su nombre a partir de observaciones de los movimientos de los planetas.

Inicios 1900s -> Evgenij E. Slutsky y George Udny Yule

- A inicios del siglo XX, Slutsky y Yule mostraron que series de tiempo con propiedades similares a series económicas pueden generarse como sumas o restas (simples o ponderadas) de procesos puramente aleatorios.
- Desarrollaron los procesos de media móvil y autorregresivos como modelos para representar series de tiempo.

1919 -> Warren M. Persons

- En 1919 Persons propuso la descomposición de una serie de tiempo en componentes no observables que dependen de causas distintas.
- Los componentes son:
 - un movimiento de largo plazo, la *tendencia*,
 - un componente cíclico con períodos de más de un año, el *ciclo económico*,
 - un componente que contiene los aumentos y disminuciones dentro del año, el ciclo *estacional*,
 - un componente que contiene todos los movimientos no atribuibles a los demás componentes, el residual o *componente irregular*.

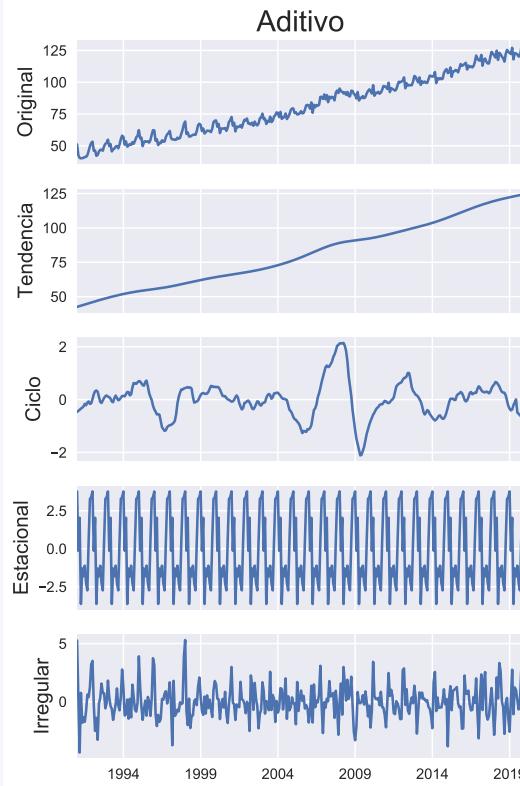
Aditiva

$$Y_t = T_t + C_t + S_t + I_t$$

Multiplicativa

$$Y_t = T_t \times C_t \times S_t \times I_t$$

Ejemplo 1-1: Componentes del IMAE



Multiplicativo

Esta es una descomposición “ingenua”. Más adelante en el curso estudiaremos métodos más sofisticados.

Enfoques clásico vs moderno de descomposición de series

- Como los componentes *no son observables*, es necesario hacer supuestos acerca de su naturaleza para estudiar el proceso generador de la serie.

Enfoque clásico	Enfoque moderno
$Y_t = \underbrace{T_t + C_t + S_t}_{\text{determinísticos}} + \underbrace{I_t}_{\text{estocástico}}$ <p>Los componentes sistemáticos son funciones <i>determinísticas</i> del tiempo, el componente residual es estocástico pero <i>no contiene movimientos sistemáticos</i>.</p>	$Y_t = \underbrace{T_t + C_t + S_t}_{\text{estocásticos}} + I_t$ <p>La “ley de movimiento” de la serie es vista como un <i>proceso estocástico</i>, y los datos de la serie como <i>una realización</i> del proceso generador de datos.</p>

1930s -> Jan Tinbergen

- En 1936, Tinbergen construyó el primer modelo econométrico: un modelo macroeconómico de la economía holandesa, empezando así el desarrollo de la econometría aplicada.
- En 1939, presentó su metodología para el análisis estadístico de teorías de ciclo económico, así como un modelo macroeconómico de Estados Unidos.
- Sus métodos fueron controversiales:

“Nadie podría ser más franco, más meticuloso, más libre de prejuicios subjetivos o partidistas que el profesor Tinbergen. No hay nadie, por lo tanto, en lo que respecta a las cualidades humanas, a quien sería más seguro confiar con magia negra. Todavía no estoy convencido de que haya alguien en quien confiaría en la etapa actual o que esta marca de alquimia estadística está madura para convertirse en una rama de la ciencia. Pero Newton, Boyle y Locke jugaron con la alquimia. Así que déjenle continuar.”

Keynes, 1940

Citado por Boumans (2015)

1938 -> Herman Wold

- Wold sistematizó y generalizó el trabajo de Slutsky y Yule.
- Demostró que para proceso covarianza-estacionario, puramente no-determinístico, existe una descomposición de la serie como combinación lineal de una serie de variables aleatorias no correlacionadas con media cero y varianza constante.

1949 -> Donald Cochrane y Guy H. Orcutt

- Tinbergen utilizó los supuestos del modelo clásico de regresión lineal, sin prestar atención a la dependencia cronológica de los residuos.
- En 1949 Cochrane y Orcutt notaron que esta práctica era problemática.
- Demostraron que si los residuos de una regresión están positivamente correlacionados, entonces la varianza de los parámetros de regresión está subestimada y los estadísticos t y F sobreestimados.
- El problema se podía resolver transformando los datos adecuadamente.

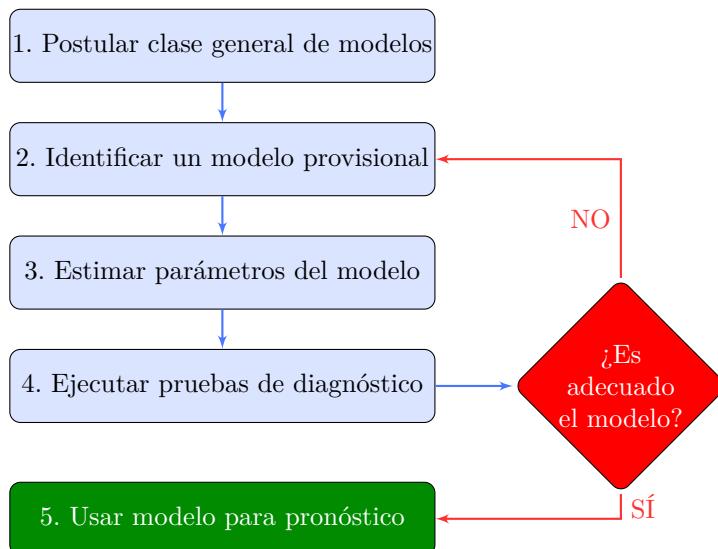
1950 -> James Durbin y Geoffrey S. Watson

- En 1950/51 Durbin y Watson desarrollaron un test para identificar autocorrelación de primer orden en los residuos.

1970 -> George E.P. Box y Gwilym M. Jenkins

- En 1970 Box y Jenkins publican un libro de texto de análisis de series de tiempo.
- Introducen modelos univariados para series de tiempo, que usan sistemáticamente la información contenida en los valores de la serie.
- Manera sencilla de producir pronósticos.
- Hoy en día conocida como *metodología Box-Jenkins*.

La metodología Box-Jenkins



1975 -> Clive W.J. Granger y Paul Newbold

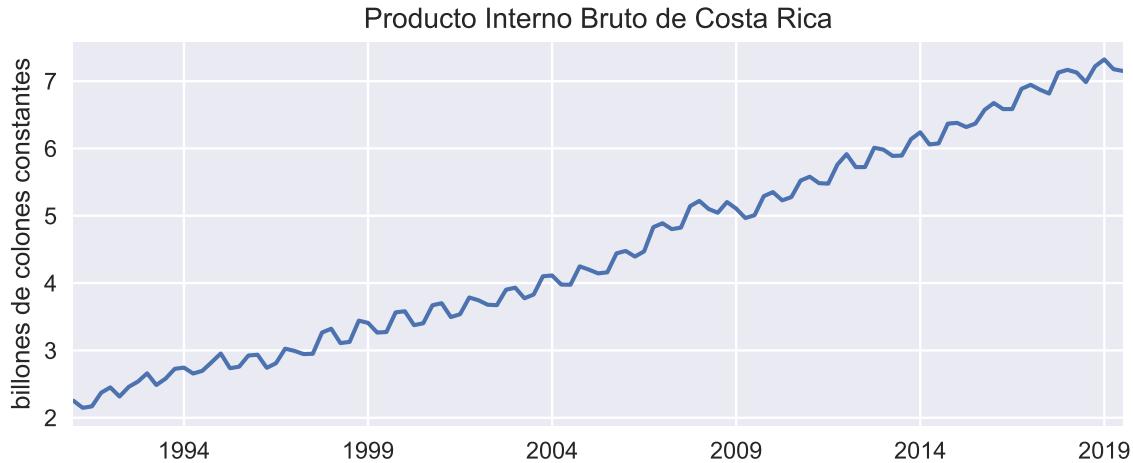
- En 1975 Granger y Newbold mostraron que los pronósticos sencillos de métodos univariados a menudo eran mejores que los pronósticos basados en modelos econométricos grandes (cientos de ecuaciones).
- Esto impulsó la enorme popularidad de la metodología Box-Jenkins.

1.2 Representación gráfica de series de tiempo

Nivel de la serie

- Antes de modelar una serie de tiempo, es útil representarla con un gráfico para detectar algunas de sus propiedades.
- En este caso: el PIB
 - muestra una tendencia positiva

- tiene variaciones estacionales
- En lo que sigue, nos referimos a esta serie en nivel como y_t .

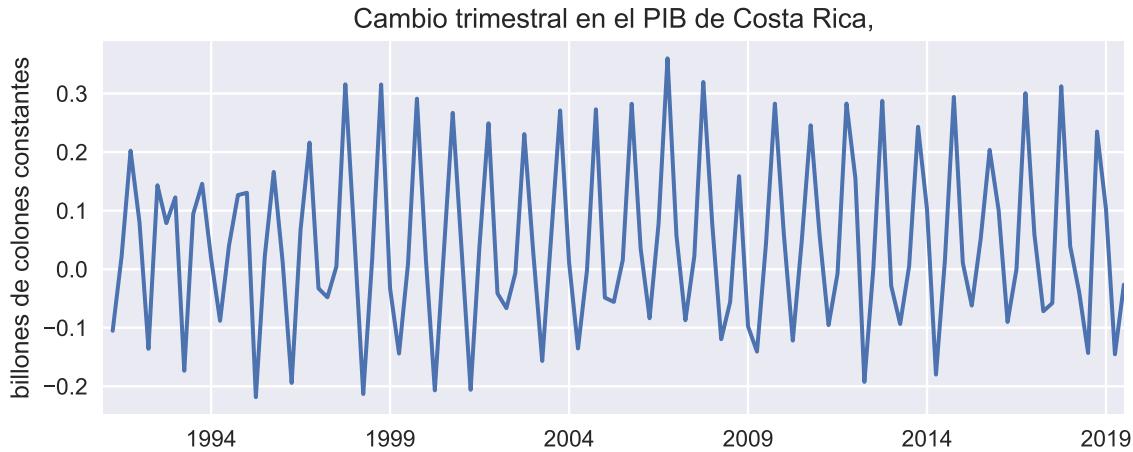


Primera diferencia de la serie

$$\Delta y_t \equiv y_t - y_{t-1}$$

Esta transformación

- elimina la tendencia de la serie,
- mantiene las oscilaciones estacionales.

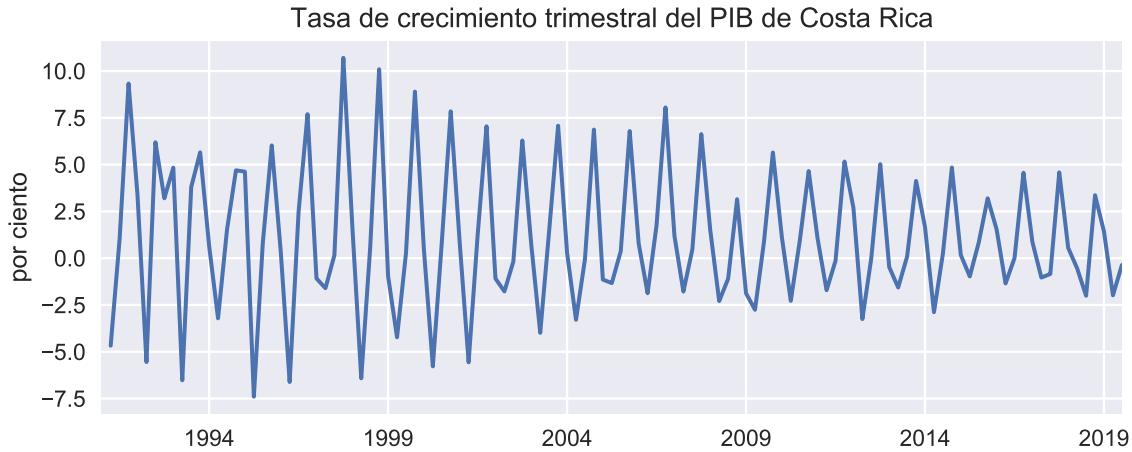


Tasa de crecimiento de la serie

$$\Delta\%y_t \equiv \frac{y_t - y_{t-1}}{y_{t-1}} \times 100$$

- Elimina tendencia, mantiene estacionalidad.

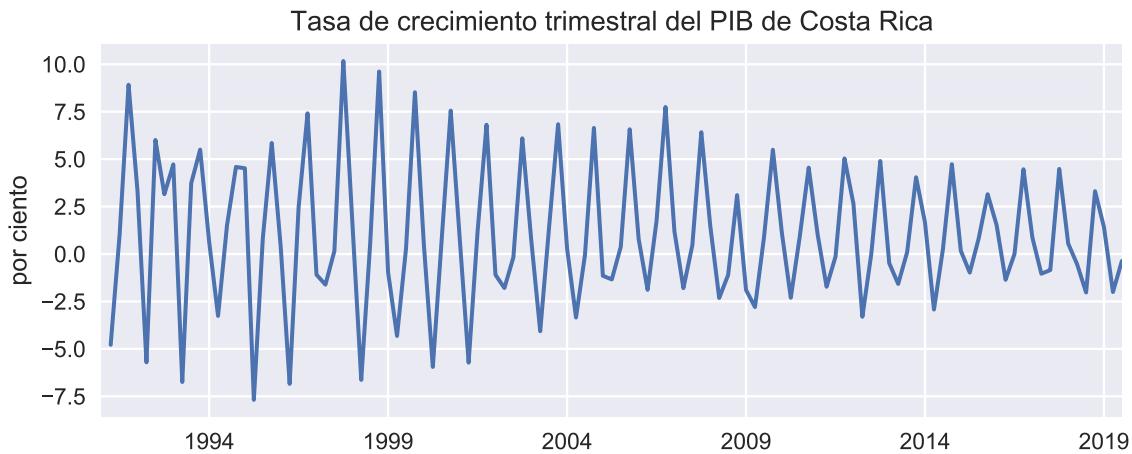
- Limitación: asimetría con respecto a cambios positivos y negativos: Subir de 100 a 125 (aumento de 25%), bajar de 125 a 100 (caída de “solo” 20%).



Tasa “continua” de crecimiento de la serie

$$\Delta\%y_t \approx (\ln y_t - \ln y_{t-1}) \times 100$$

- Similar a la anterior porque $\ln(1 + x) \approx x$ si x es “pequeño”
- Ventaja: simetría con respecto a cambios positivos y negativos



Diferencia interanual de la serie

$$\Delta_4 y_t \equiv y_t - y_{t-4}$$

- Elimina tanto la tendencia como el componente estacional
- Nótese la fuerte disminución del PIB durante la crisis de 2008.

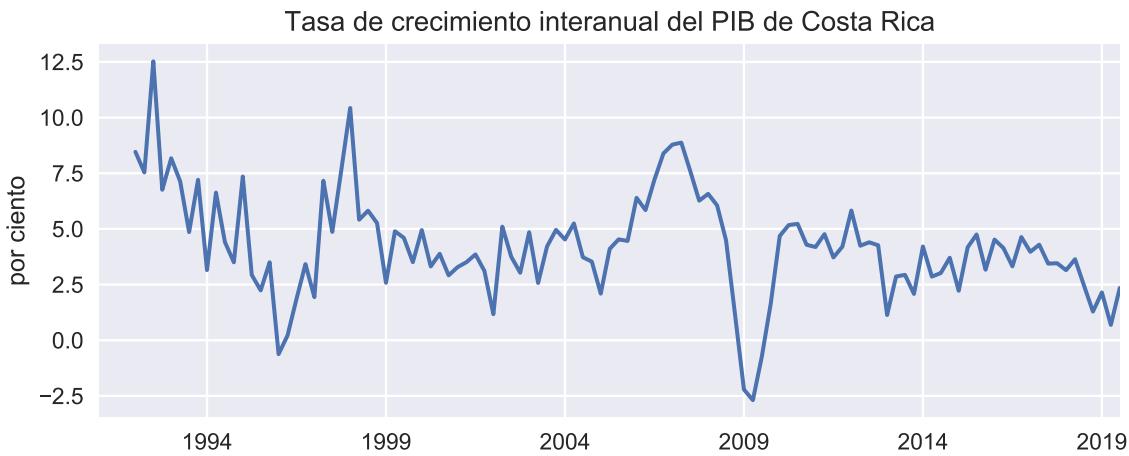


Tasa de crecimiento interanual

$$\Delta_4\%y_t \approx (\ln y_t - \ln y_{t-4}) \times 100$$

- Equivalente a la suma de las tasas de crecimiento de los cuatro trimestres comprendidos en el año:

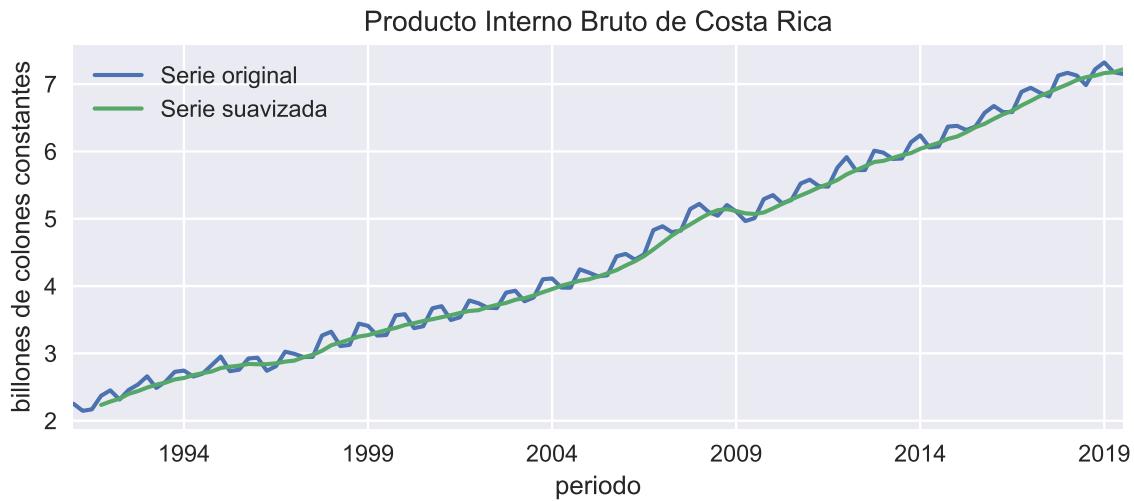
$$\begin{aligned} \Delta_4\%y_t &\approx (\ln y_t - \ln y_{t-4}) \times 100 \\ &= (\ln y_t - \ln y_{t-1} + \ln y_{t-1} - \ln y_{t-2} + \ln y_{t-2} - \ln y_{t-3} + \ln y_{t-3} - \ln y_{t-4}) \times 100 \\ &= \Delta\%y_t + \Delta\%y_{t-1} + \Delta\%y_{t-2} + \Delta\%y_{t-3} \end{aligned}$$



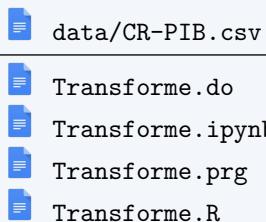
Serie suavizada por media móvil

$$y_t^s \equiv \frac{1}{4} (y_t + y_{t-1} + y_{t-2} + y_{t-3})$$

- Elimina el componente estacional, pero manteniendo la tendencia
- Se observa un cambio estructural en 2008-2009.



Ejemplo 1-2: Transformación de datos



Las transformaciones ilustradas en las figuras pueden ser calculadas con Stata y con Python:

	Stata	Python
Serie original	pib	pib
Primera diferencia	D.pib	pib.diff()
Tasa de crecimiento	D.pib / L.pib	pib.pct_change(1)
Tasa de variación continua	gen l=log(pib) D.lpib	np.log(pib).diff()
Diferencia interanual	S4.pib	pib.diff(4)
Tasa de crecimiento interanual	S4.lpib	np.log(pib).diff(4)
Suavizada por media móvil	tssmooth ma y = pib, window(3 1 0)	pib.rolling(4).mean()

1.3 El operador de rezagos

Definición de operador de series de tiempo

- Un operador de serie de tiempo es un “proceso” que transforma una o más series de tiempo en nuevas series de tiempo.
- Ejemplos:
 - Multiplicación escalar: $y_t = \beta x_t$
 - Suma: $y_t = x_t + w_t$
 - Identidad: $y_t = 1y_t$
- Nótese que:

$$y_t = \beta(x_t + w_t) = \beta x_t + \beta w_t$$

Operador de rezago

- El operador de rezago se denota por L y se define como:

$$L x_t \equiv x_{t-1}$$

- En general, se tiene que:

$$L^k x_t = x_{t-k}$$

Ejemplo 1-3: Operador de rezagos y transformación de series

Algunas de las transformaciones de la serie y_t de la sección anterior pueden expresarse con el operador de rezagos:

Serie original	y_t
Primera diferencia	$\Delta y_t \equiv y_t - y_{t-1} = y_t - L y_t = (1 - L)y_t$
Tasa de crecimiento	$\Delta \% y_t \approx 100 (\ln y_t - \ln y_{t-1}) = 100(1 - L) \ln y_t$
Diferencia interanual	$\Delta_4 y_t \equiv y_t - y_{t-4} = y_t - L^4 y_t = (1 - L^4)y_t$
Tasa de crecimiento interanual	$\Delta_4 \% y_t \approx 100 (\ln y_t - \ln y_{t-4}) = 100(1 - L^4) \ln y_t$
Suavizada por media móvil	$y_t^s \equiv \frac{1}{4} (y_t + y_{t-1} + y_{t-2} + y_{t-3}) = \frac{1}{4} (1 + L + L^2 + L^3) y_t$

Ejemplo 1-4: Operador de rezagos

 data/LandD.csv

 LandD.do
 LandD.ipynb
 landd.prg
 LandD.R

t	y_t	$L y_t$	$L^2 y_t$	Δy_t	$\Delta^2 y_t$	$\Delta_4 y_t$
2018Q1	10
2018Q2	13	10	.	3	.	.
2018Q3	10	13	10	-3	-6	.
2018Q4	8	10	13	-2	1	.
2019Q1	15	8	10	7	9	5
2019Q2	16	15	8	1	-6	3
2019Q3	14	16	15	-2	-3	4
2019Q4	11	14	16	-3	-1	3

Propiedades del operador de rezago

Sean x_t, w_t dos series de tiempo. Entonces:

- $L(\beta x_t) = \beta L x_t$
- $L(x_t + w_t) = L x_t + L w_t$
- $L(c) = c$
- $L^{-h} x_t = x_{t+h}$
- $L^0 x_t = x_t$
- $(\alpha L^h + \beta L^k)x_t = \alpha x_{t-h} + \beta x_{t-k}$

donde α, β, c son constantes.

Polinomio de rezagos

- El operador de rezagos sigue las reglas usuales de operaciones algebraicas. Por ejemplo:

$$\begin{aligned}
 (a + bL)(c + dL)x_t &= (a + bL)(cx_t + dx_{t-1}) \\
 &= acx_t + adx_{t-1} + bcx_{t-1} + bd़x_{t-2} \\
 &= [ac + (ad + bc)L + bdL^2] x_t
 \end{aligned}$$

- Así, definimos un polinomio de rezagos de orden p :

$$(1 + \phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p) x_t = x_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p}$$

Inverso de un polinomio de rezagos de grado 1

- Considere la operación

$$(1 - \phi L) (1 + \phi L + \cdots + \phi^k L^k) x_t = (1 - \phi^{k+1} L^{k+1}) x_t \\ = x_t - \phi^{k+1} x_{t-k-1}$$

- Si $|\phi| < 1$,

$$\lim_{k \rightarrow \infty} \phi^{k+1} x_{t-k-1} = 0$$

- con lo que

$$(1 - \phi L) (1 + \phi L + \phi^2 L^2 + \dots) x_t = x_t$$

- En este caso, escribimos

$$(1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \dots$$

Inverso de un polinomio de rezagos de grado p

- Consideremos el polinomio

$$\Phi(L) \equiv 1 - \phi_1 L - \cdots - \phi_p L^p$$

- Si factorizamos el polinomio como

$$\Phi(L) = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L)$$

- Encontramos su inverso como

$$\Phi^{-1}(L) = (1 - \lambda_1 L)^{-1} \cdots (1 - \lambda_p L)^{-1} \\ = (1 + \lambda_1 L + \lambda_1^2 L^2 + \dots) \cdots (1 + \lambda_p L + \lambda_p^2 L^2 + \dots)$$

1.4 Procesos estocásticos

Procesos estocásticos

- Un *proceso estocástico* es una secuencia temporal de variables aleatorias $\{Y_t\}_{t=-\infty}^{\infty}$.
- Dos tipos de procesos:

Continuos si sus realizaciones son tomadas de un intervalo de la recta real $Y_t \in [a, b] \subseteq \mathbb{R}$.

Discretos si hay solamente un número contable de realizaciones $Y_t \in \{y_1, y_2, \dots, y_n\}$.

- También llamado *proceso generador de datos*.

Procesos estocásticos i.i.d.

- Los elementos de un proceso estocástico son *idéntica e independientemente distribuidos* (*abreviado “iid”*), si la distribución de probabilidad es la misma para cada miembro del proceso Z_t e independiente de las realizaciones de otros miembros del proceso.
- En este caso, para la muestra $\{Y_t\}_{t=1}^T$:

$$\mathbb{P}[Y_1 = y_1, Y_2 = y_2, \dots, Y_T = y_T] = \\ \mathbb{P}(Y_1 = y_1) \times \mathbb{P}(Y_2 = y_2) \times \cdots \times \mathbb{P}(Y_T = y_T)$$

Momentos incondicionales

- Función de distribución acumulada incondicional

$$F_{Y_t}(y) = \mathbb{P}[Y_t \leq y]$$

- Esperanza (media) incondicional

$$\mu_t \equiv \mathbb{E}(Y_t) = \int_{-\infty}^{\infty} y \, dF_{Y_t}(y)$$

- Varianza incondicional

$$\gamma_{0t} \equiv \mathbb{E}(Y_t - \mu_t)^2 = \int_{-\infty}^{\infty} (y - \mu_t)^2 \, dF_{Y_t}(y)$$

- Autocovarianza

$$\gamma_{jt} \equiv \mathbb{E}(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})$$

Estacionariedad

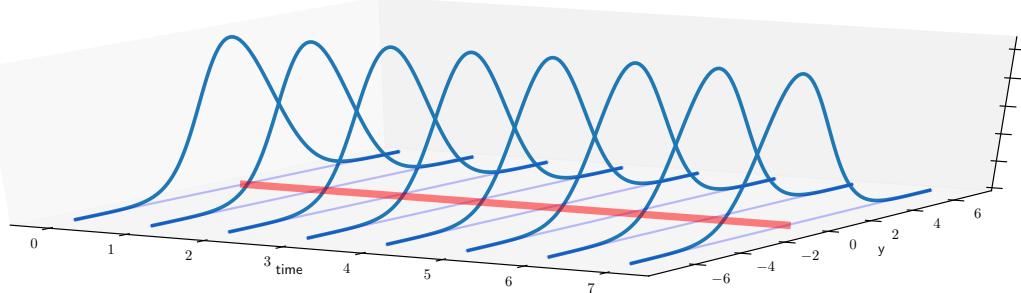
Si la media μ_t ni las autocovarianzas γ_{jt} dependen de la fecha t , entonces decimos que el proceso Y_t es *covarianza-estacionario* o *débilmente estacionario*:

$$\begin{aligned} \mathbb{E}(Y_t) &= \mu && \text{para todo } t \\ \mathbb{E}(Y_t - \mu)(Y_{t-j} - \mu) &= \gamma_j && \text{para todo } t \text{ y cualquier } j \end{aligned}$$

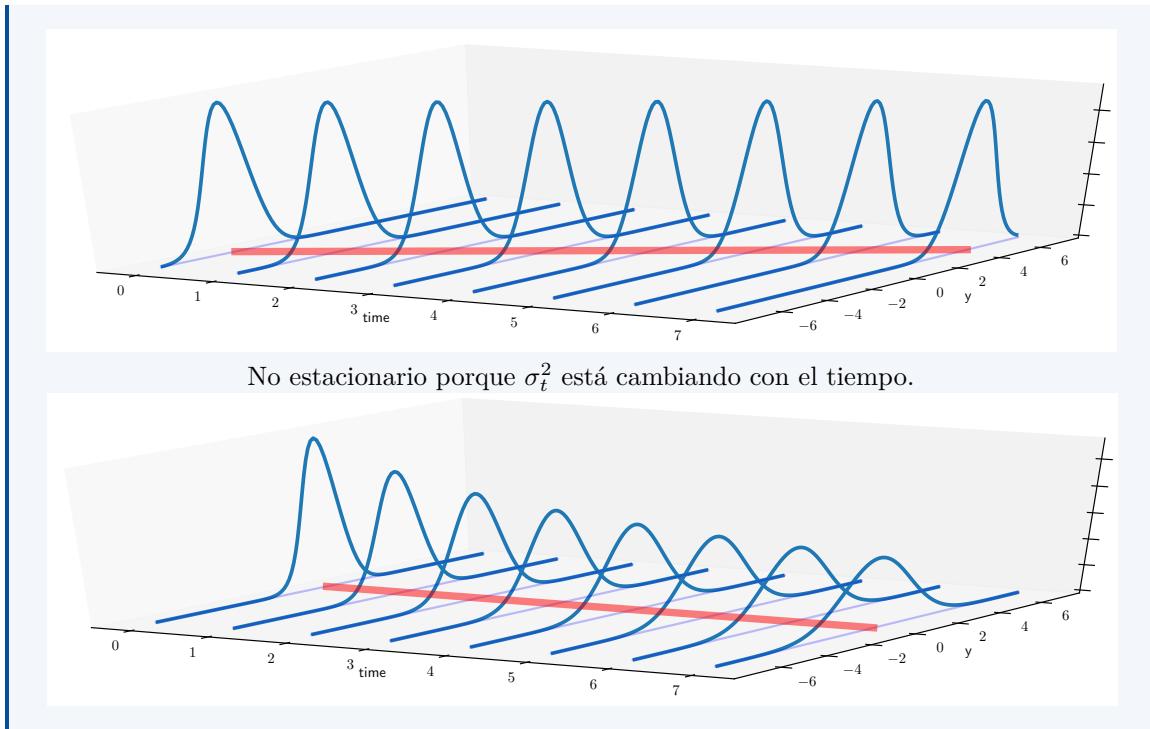
Ejemplo 1-5: Procesos estacionarios y no estacionarios

Supongamos que Y_t es un proceso estocástico tal que $Y_t \sim N(\mu_t, \sigma_t^2)$

Estacionario porque μ_t y σ_t^2 son constantes.



No estacionario porque μ_t está cambiando con el tiempo.



Ruido blanco

- El bloque básico para construir los procesos considerados en este curso es una secuencia $\{\epsilon_t\}$ cuyos elementos tienen media cero y varianza σ^2 ,

$$\mathbb{E}(\epsilon_t) = 0 \quad (\text{media cero})$$

$$\mathbb{E}(\epsilon_t^2) = \sigma^2 \quad (\text{varianza constante})$$

$$\mathbb{E}(\epsilon_t \epsilon_\tau) = 0 \quad \text{for } t \neq \tau \quad (\text{términos no correlacionados})$$

- Si los términos están normalmente distribuidos

$$\epsilon_t \sim N(0, \sigma^2)$$

entonces tenemos el *proceso ruido blaco gaussiano*.

1.5 Series de tiempo

Definición de serie de tiempo

- Una serie de tiempo es una colección de observaciones indexadas por la fecha de cada observación, denotada por y_t

$$\{y_1, y_2, \dots, y_T\}$$

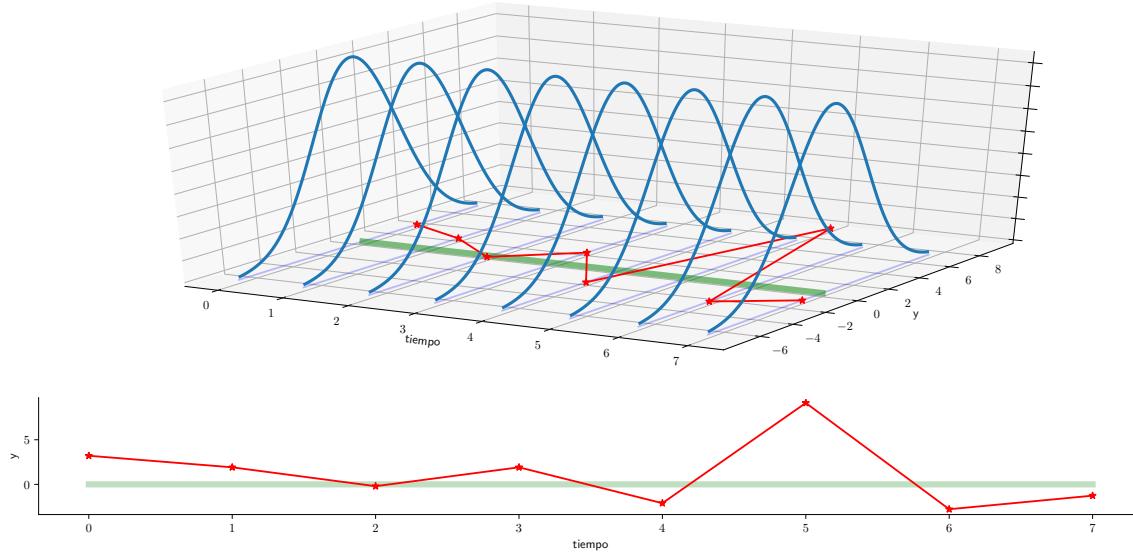
- En la práctica se asume que lo anterior es sólo una muestra, pero que la serie pudo haber sido observada en más períodos.

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, y_2, \dots, y_T, y_{T+1}, y_{T+2}, \dots\}$$

- Así, vemos a una serie de tiempo como *una realización* del proceso estocástico $\{Y_t\}_{t=-\infty}^{\infty}$, de la cual solo tenemos observaciones para t entre 1 y T .

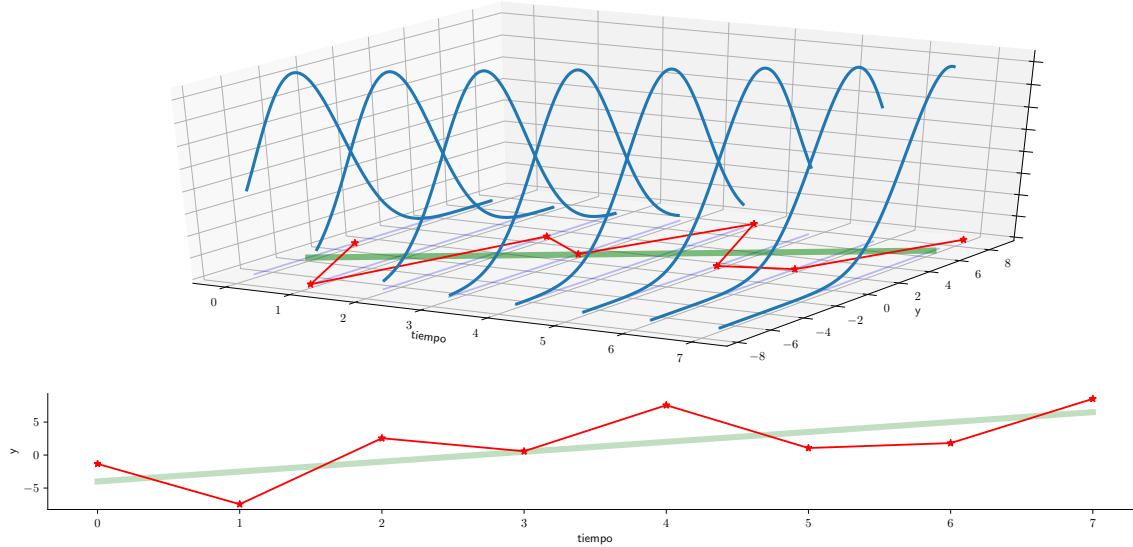
Serie de tiempo estacionaria

Una serie de tiempo estacionaria es una realización de un proceso estocástico estacionario.



Serie de tiempo no estacionaria

Una serie de tiempo no estacionaria es una realización de un proceso estocástico no estacionario.



Casos particulares de series de tiempo

- Tendencia:

$$y_t = t$$

- Constante:

$$y_t = c$$

- Proceso ruido blanco de Gauss:

$$y_t = \epsilon_t, \text{ con } \{\epsilon_t\}_{t=-\infty}^{\infty} \text{ i.i.d., } \epsilon_t \sim N(0, \sigma^2)$$

Nota: Ley de los grandes números

La ley débil de los grandes números establece que si X_1, X_2, X_3, \dots es una sucesión infinita de variables aleatorias *independientes* que tienen el mismo valor esperado μ y varianza σ^2 , entonces el promedio

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

converge en probabilidad a μ . En otras palabras, para cualquier número positivo ϵ se tiene

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| < \epsilon) = 1.$$

Prueba (muy) informal: Note que:

- $\mathbb{E}[\bar{X}_n] = \mu$
- $\text{Var}[\bar{X}_n] = \frac{\sigma^2}{n} \rightarrow 0$ conforme $n \rightarrow \infty$.

Infiriendo los momentos del proceso estocástico

- Suponga que tenemos un número grande n de realizaciones independientes del proceso estocástico $\{Y_t\}_{t=1}^T$, y que deseamos estimar el valor del momento no central $\mu'_{k,t} \equiv \mathbb{E}[Y_t^k]$.
- Utilizando la ley de los grandes números, podemos estimarlo así

$$\hat{\mu}'_{k,t} = \frac{y_{t,1}^k + y_{t,2}^k + \dots + y_{t,n}^k}{n}$$

- En la práctica, sólo tenemos una única realización del proceso: y_1, y_2, \dots, y_T .
- ¿Podemos estimar $\mu'_{k,t}$ a partir de $\bar{y}^k = \frac{y_1^k + y_2^k + \dots + y_T^k}{T}$?
- En general no, a menos que el proceso estocástico sea
 1. estacionario
 2. ergódico (para el momento k).

Ergodicidad

En el caso de procesos estacionarios, el supuesto de *ergodicidad* significa que momentos muestrales calculados a partir de una serie de tiempo con un número finito de observaciones converge (en algún sentido) a sus contrapartes poblacionales.

Ergódico en media

El proceso es ergódico en media si

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=1}^T y_t - \mu \right)^2 \right] = 0$$

Ergódico en varianza

El proceso es ergódico en varianza si

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=1}^T (y_t - \mu)^2 - \sigma_y^2 \right)^2 \right] = 0$$

Algunas apuntes acerca de la ergodicidad

- Las condiciones de ergodicidad son “propiedades de consistencia” para variables aleatorias dependientes, y no pueden ser evaluadas. Por ello, deben ser asumidas.
- No podemos usar la ley de los grandes números para probar ergodicidad, porque las distribuciones de distintos momentos t pueden ser idénticas, pero en general no son independientes.
- Un proceso no estacionario no puede ser ergódico: ¿a los momentos de cuál variable Y_t convergerían los momentos muestrales si los poblacionales no fueran constantes?

Estimación de momentos de un proceso estacionario

Si el proceso es ergódico, podemos estimar consistentemente:

Media

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

Varianza

$$\hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2$$

Autocovarianza

$$\hat{\gamma}_\tau = \frac{1}{T} \sum_{t=1}^{T-\tau} (y_t - \hat{\mu})(y_{t+\tau} - \hat{\mu})$$

Autocorrelación

$$\hat{\rho}_\tau = \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0}$$

para $\tau = 1, 2, \dots, T-1$.

Asimetría

$$\hat{S} = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^3}{\sqrt{\hat{\gamma}_0^3}}$$

Kurtosis

$$\hat{K} = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^4}{\hat{\gamma}_0^2}$$

Autocorrelación

- Es imposible evaluar qué tan fuerte es la dependencia de variables de un proceso estocástico a partir de las autocovarianzas, porque éstas dependen de las unidades de medida de la serie.
- Por ello, utilizamos las *autocorrelaciones*, que se obtienen de dividir las autocovarianzas por la varianza del proceso.
- Si el proceso es estacionario se cumple que:
 - $\rho_0 = 1$
 - $\rho_\tau = \rho_{-\tau}$
 - $|\rho_\tau| \leq 1$ para todo τ .
- Llamamos *autocorrelograma* a la secuencia de autocorrelaciones ρ_τ , vista como función del número de rezagos τ .

Autocorrelación muestral

- En la práctica, a partir de una muestra estimaremos la *autocorrelación muestral*.
- Como nuestra muestra es finita, los valores estimados no necesariamente coincidirán con los poblacionales.
- Por ello, nos interesa especialmente saber cuándo estas autocorrelaciones son significativamente distintas de cero.
- Según Bartlett (1946), la varianza de los coeficientes de autocorrelación en los cuales $\rho_{k+1}, \rho_{k+2}, \dots = 0$ está dada por

$$V[\hat{\rho}(\tau)] \approx \frac{1}{T} \left(1 + 2 \sum_{j=1}^k \rho(j)^2 \right)$$

Autocorrelación parcial

- La m -ésima autocorrelación parcial $\hat{a}_m^{(m)}$ se estima por OLS como el último coeficiente de una regresión de y en una constante y sus m valores más recientes:

$$y_t = c + \hat{a}_1^{(m)} y_{t-1} + \hat{a}_2^{(m)} y_{t-2} + \dots + \hat{a}_m^{(m)} y_{t-m} + \epsilon_t \hat{\epsilon}_{t+1}$$

- La idea es medir el efecto *directo* de, por ejemplo, y_{t-1} sobre y_{t+1} , controlando por el efecto indirecto que y_{t-1} pueda tener sobre y_{t+1} a través de y_t .
- Si los datos están generados por un proceso AR(p), entonces el estimador $\hat{a}_m^{(m)}$ tiene una varianza alrededor del verdadero valor (cero) aproximada por

$$\text{Var}[\hat{a}_m^{(m)}] \approx \frac{1}{T} \quad \text{para } m = p+1, p+2, \dots$$

Ejemplo 1-6: Autocorrelograma del IMAE



log_imae.csv



correlogram.ipynb



correlogram.do

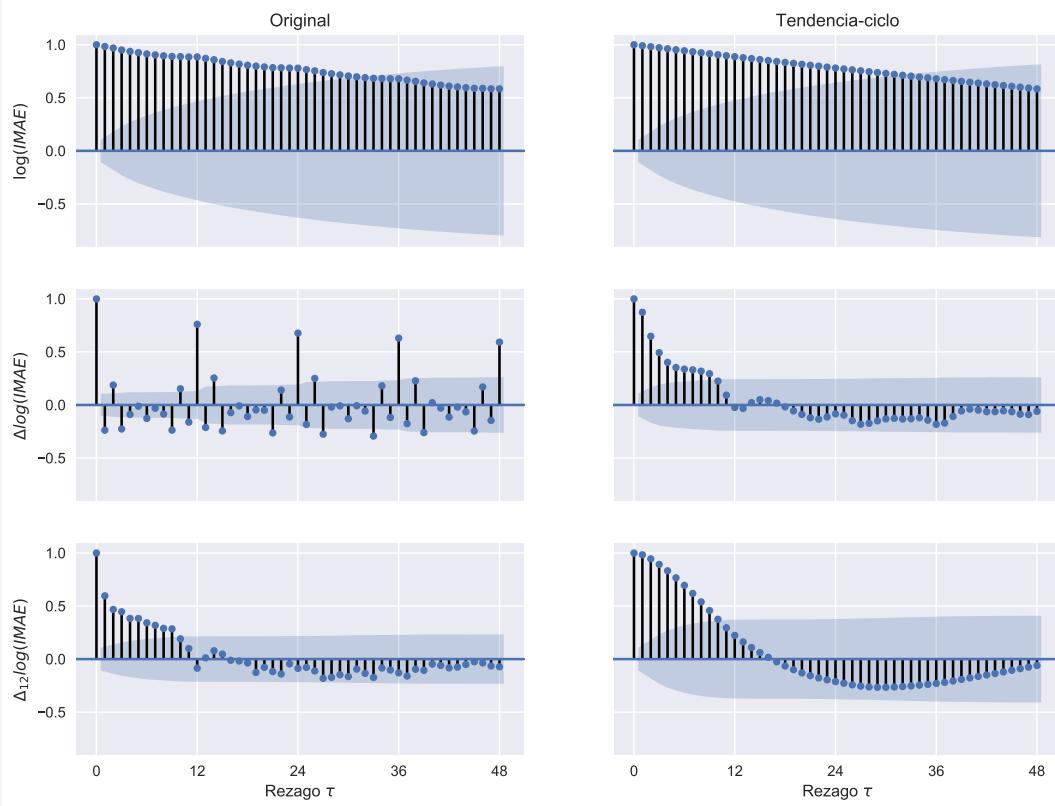


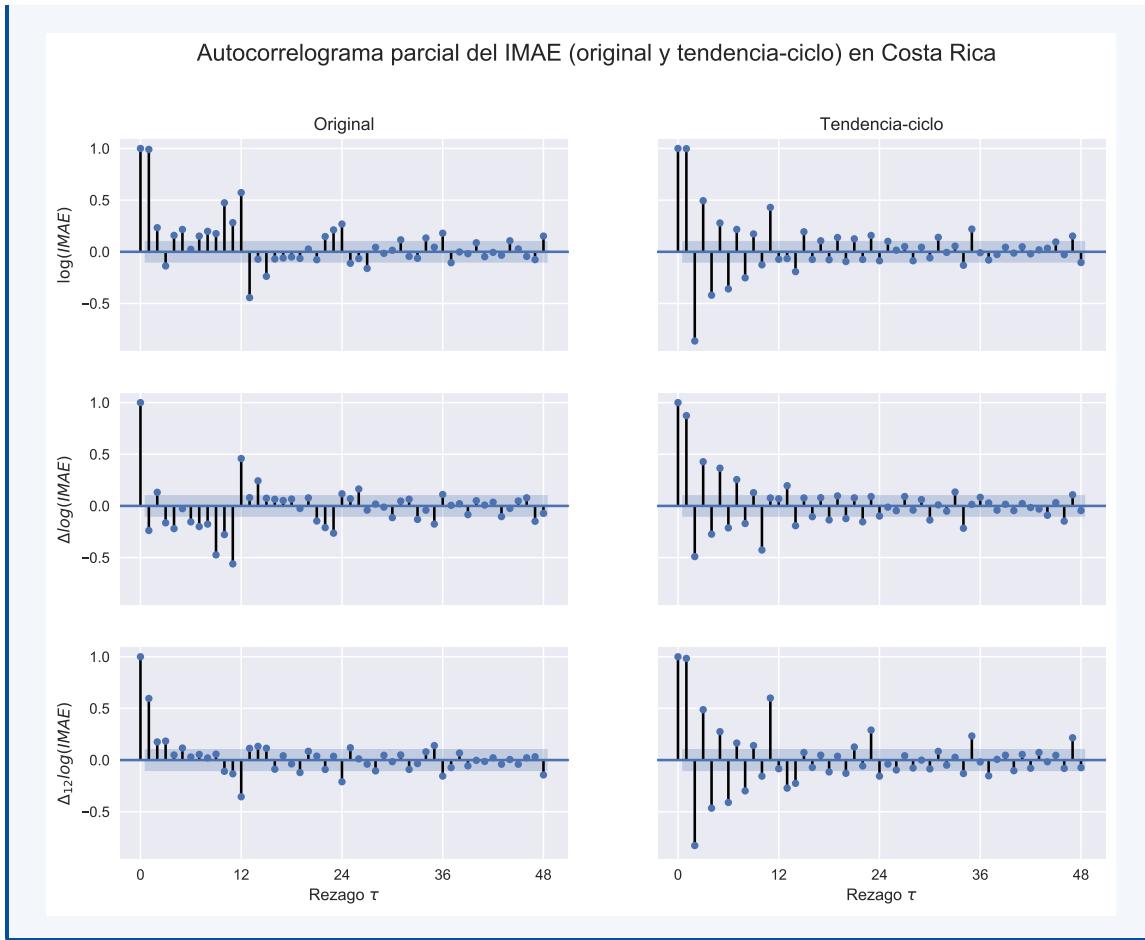
correlogram.prg



correlogram.R

Autocorrelograma del IMAE (original y tendencia-ciclo) en Costa Rica





Determinando si una serie de tiempo es ruido blanco

- Cuando se estiman modelos de series de tiempo, es importante evaluar si los residuos de la estimación corresponden a una realización de un proceso de ruido blanco.
- Recordando que un proceso ruido blanco $\{\epsilon_t\}$ es tal que

$$\mathbb{E}(\epsilon_t) = 0 \quad (\text{media cero})$$

$$\mathbb{E}(\epsilon_t^2) = \sigma^2 \quad (\text{varianza constante})$$

$$\mathbb{E}(\epsilon_t \epsilon_\tau) = 0 \quad \text{for } t \neq \tau \quad (\text{términos no correlacionados})$$

una forma natural de evaluar si los residuos son ruido blanco es determinar si las autocorrelaciones

$$\rho_1 = \rho_2 = \dots = \rho_\tau = 0$$

para todo $\tau \geq 1$

Test de Box-Pierce

Test de Box-Pierce (1970)

¿Es esta serie un caso de ruido blanco?

H₀ $\rho_1 = \rho_2 = \dots = \rho_m = 0$ (sí es ruido blanco)**Test**

$$Q^* = T \sum_{j=1}^m \hat{\rho}_j^2 \stackrel{\text{asy}}{\sim} \chi_{m-k}^2$$

k es número de parámetros estimados

Si $Q^* > \chi_{m-k}(1 - \alpha)$, rechazar H_0 con 100% de significancia: la serie no es ruido blanco.

La intuición es que si la serie no es ruido blanco, algunos $\hat{\rho}_j$ serán “muy grandes”, y entonces Q^* también lo será.

Test de Ljung-Box**Test de Ljung-Box (1978)**

¿Es esta serie un caso de ruido blanco?

H₀ $\rho_1 = \rho_2 = \dots = \rho_m = 0$ (sí es ruido blanco)**Test**

$$Q = T(T+2) \sum_{j=1}^m \frac{\hat{\rho}_j^2}{T-j} \stackrel{\text{asy}}{\sim} \chi_{m-k}^2$$

Si $Q > \chi_{m-k}(1 - \alpha)$, rechazar H_0 con 100% de significancia: la serie no es ruido blanco.

Este test es similar al de Box-Pierce, pero ajustada para muestras pequeñas.

Test de Durbin-Watson**Test de Durbin-Watson (1950/1)**

¿Hay autocorrelación de primer orden en esta serie?

H₀ $\rho_1 = 0$ (no hay autocorrelación)**Test**

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \approx 2(1 - \hat{\rho}_1) \quad (\text{si } T \text{ es grande})$$

Si d está “lejos” de 2 según los valores críticos de DW, rechazar H_0 : la serie sí presenta autocorrelación.

Esta prueba no es válida para residuos de una ecuación donde haya rezagos de la variable dependiente.

Ventajas de Box-Pierce / Ljung-Box sobre Durbin-Watson

1. Box-Pierce / Ljung-Box evalúan la existencia de autocorrelación de cualquier orden, no solo de primer orden.
2. Sus resultados también son válidos para evaluar residuos de regresiones que contienen rezagos de la variable dependiente.

Ejemplo 1-7: Pruebas de ruido blanco

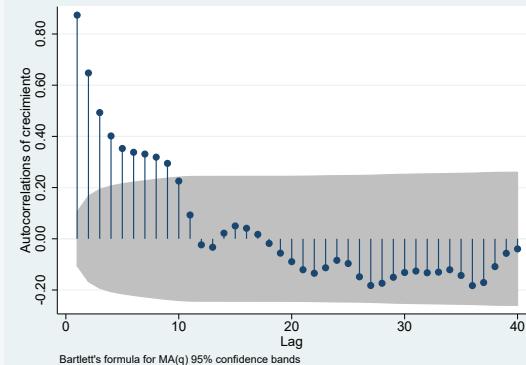
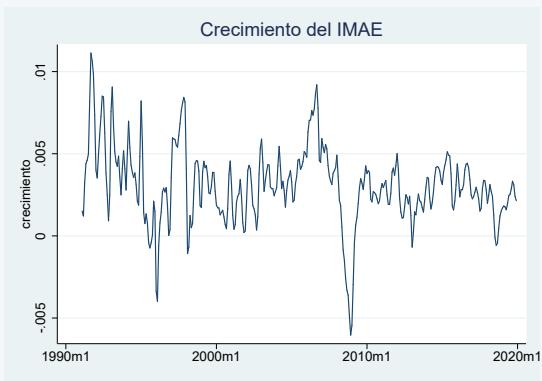
log_imae.csv
 euro.csv

 wntest.ipynb
 wntest.do

Crecimiento del IMAE

Los resultados de las pruebas Ljung-Box son consistentes con lo que obtuvimos a partir de un autocorrelograma: el crecimiento mensual del IMAE no es ruido blanco.

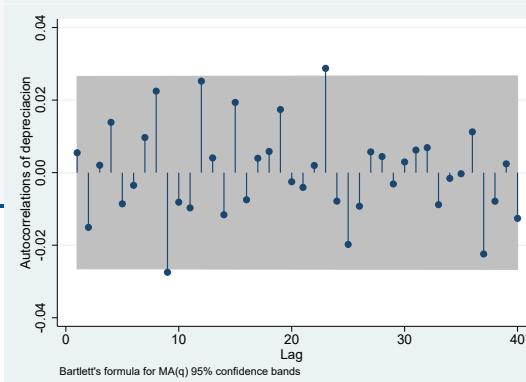
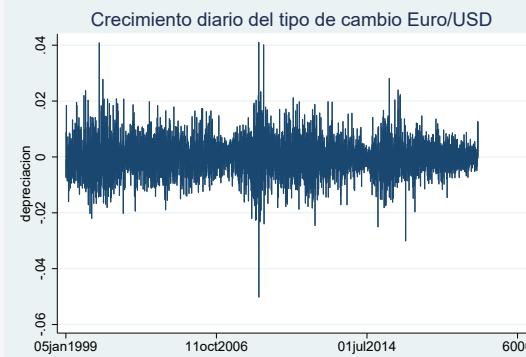
LAG	AC	PAC	Q	Prob>Q
1	0.8737	0.8741	267.19	0.0000
2	0.6478	-0.4903	414.48	0.0000
3	0.4931	0.4280	500.08	0.0000
4	0.4018	-0.2733	557.07	0.0000
5	0.3528	0.3654	601.15	0.0000
6	0.3378	-0.2128	641.68	0.0000
7	0.3310	0.2552	680.7	0.0000
8	0.3190	-0.1704	717.05	0.0000
9	0.2947	0.1289	748.16	0.0000
10	0.2256	-0.4269	766.44	0.0000
11	0.0931	0.0771	769.57	0.0000
12	-0.0235	0.0692	769.77	0.0000



Cambios en el tipo EUR/USD

Las pruebas Ljung-Box *no rechazan* que esta serie sea ruido blanco. Pero en la gráfica parece que la varianza no es constante, por lo que posiblemente la serie tampoco sería ruido blanco.

LAG	AC	PAC	Q	Prob>Q
1	0.0055	0.0055	.16483	0.6847
2	-0.0151	-0.0151	1.4129	0.4934
3	0.0020	0.0022	1.4357	0.6972
4	0.0139	0.0136	2.4911	0.6462
5	-0.0086	-0.0087	2.8965	0.7159
6	-0.0035	-0.0030	2.9636	0.8134
7	0.0097	0.0094	3.4785	0.8375
8	0.0225	0.0222	6.2474	0.6195
9	-0.0275	-0.0273	10.384	0.3203
10	-0.0081	-0.0072	10.748	0.3775
11	-0.0097	-0.0109	11.266	0.4213
12	0.0252	0.0249	14.762	0.2547



De no-correlación a independencia y normalidad

- El hecho de que una serie no esté autocorrelacionada no implica que sus elementos sean independientes o que estén normalmente distribuidos.
- Ausencia de autocorrelación implica independencia solamente si las variables están normalmente distribuidas.
- Usualmente se asume normalidad del proceso estocástico, porque muchos tests dependen de este supuesto.
- Para evaluar si este supuesto es apropiado, analizamos los momentos tercero (asimetría) y cuarto (kurtoosis).

Test de normalidad

Test de Jarque-Bera(1980)

	¿Es esta serie normal?
H₀	$S \equiv \mathbb{E} \left(\frac{y-\mu}{\sigma} \right)^3 = 0, K \equiv \mathbb{E} \left(\frac{y-\mu}{\sigma} \right)^4 = 3$ (sí es)
Test	$JB = \frac{T}{6} \left(\hat{S}^2 + \frac{1}{4} \left(\hat{K} - 3 \right)^2 \right) \sim \chi_2^2$
	Si $JB > \chi_2^2(1 - \alpha)$, rechazar H_0 con $100\alpha\%$ de significancia: la serie no es normal.

Referencias del capítulo 1

- Boumans, Marcel J. (15 de feb. de 2015). *Econometrics: The Keynes-Tinbergen Controversy*. URL: <https://dx.doi.org/10.2139/ssrn.2565355>.
- Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
- Kirchgässner, Gebhard, Jürgen Wolters y Uwe Hassler (2013). *Introduction to Modern Time Series Analysis*. 2^a ed. Springer. ISBN: 978-3-642-33435-1.
- Valadkhani, Abbas (2004). “History of macroeconometric modelling: lessons from past experience”. En: *Journal of Policy Modeling* 26, págs. 265-281.

Capítulo 2

Ecuaciones en diferencia

En este capítulo

2.1 Solución por sustituciones recursivas	26
2.2 Solución por combinación de soluciones homogéneas y particulares	33
2.3 Solución por medio del operador de rezagos	36

Introducción

- Este tema constituye el primer paso para el estudio de la econometría de series de tiempo
- En esta presentación, se asumirá que todas las variables son determinísticas (no estocásticas).
- El interés de esta presentación es el estudio de las consecuencias *dinámicas* de eventos a través del *tiempo*.
- Se limita la exposición a ecuaciones en diferencia lineales.

Ecuación en diferencia lineal de orden p

- La variable y_t evoluciona como un ecuación en diferencia de orden p cuando depende de sus últimos p valores

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

- Si $w_t = 0$ en todo período t , obtenemos la ecuación en diferencia lineal homogénea

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = 0$$

- Nuestra meta es responder a: *¿cuál es el efecto sobre la trayectoria de y de un cambio en w ?*

Ecuación en diferencia de primer orden

- Si $p = 1$, la variable y_t evoluciona como un ecuación en diferencia de primer orden

$$y_t = \phi y_{t-1} + w_t$$

- En este caso, la ecuación homogénea correspondiente es

$$y_t - \phi y_{t-1} = 0$$

2.1 Solución por sustituciones recursivas

Solución de la ecuación de primer orden

- Dado un valor inicial y_{-1} y la secuencia

$$\{w_0, w_1, \dots, w_t\}$$

la ecuación puede resolverse de manera recursiva como:

$$y_t = \phi^{t+1}y_{-1} + \phi^t w_0 + \phi^{t-1} w_1 + \dots + \phi w_{t-1} + w_t$$

Multiplicador dinámico: shock transitorio

- La solución es similar si se desea expresar y_{t+j} a partir de y_t

$$y_{t+j} = \phi^{j+1}y_{t-1} + \phi^j w_t + \phi^{j-1} w_{t+1} + \dots + \phi w_{t+j-1} + w_{t+j}$$

- El multiplicador dinámico se obtiene simplemente como:

$$\frac{\partial y_{t+j}}{\partial w_t} = \phi^j$$

- El proceso es estable si y sólo si $|\phi| < 1$

Valor presente

- Sea β el factor de descuento. Se define el valor presente:

$$VP = \sum_{j=0}^{\infty} \beta^j y_{t+j}$$

- ¿Cuál es el efecto de un cambio en w_t sobre el VP de y ?

$$\frac{\partial VP}{\partial w_t} = \sum_{j=0}^{\infty} \beta^j \frac{\partial y_{t+j}}{\partial w_t} = \sum_{j=0}^{\infty} \beta^j \phi^j = \frac{1}{1 - \beta\phi}$$

siempre y cuando $|\beta\phi| < 1$.

Efecto de un shock permanente

- Supóngase que el cambio en w_t es permanente.
- ¿Qué efecto tiene sobre y en el largo plazo?

$$\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+j}}{\partial w_{t+k}} = \lim_{j \rightarrow \infty} \sum_{k=0}^j \phi^{j-k} = \frac{1}{1 - \phi}$$

siempre y cuando $|\phi| < 1$.

Efecto acumulado de un shock transitorio

- Se desea la suma de los cambios en y como consecuencia de un único cambio en w_t .
- Esto corresponde al ejemplo del VP cuando $\beta = 1$:

$$\text{EA} = \sum_{j=0}^{\infty} 1^j \frac{\partial y_{t+j}}{\partial w_t} = \sum_{j=0}^{\infty} \phi^j = \frac{1}{1 - \phi}, \quad |\phi| < 1.$$

Nótese que el efecto acumulado de un shock transitorio es igual al efecto de un shock permanente en el largo plazo.

Ecuación en diferencia de orden p

- La variable y_t evoluciona como una ecuación en diferencia de primer orden cuando depende de sus últimos p valores

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

- Es muy complicado analizar por sustitución recursiva la dinámica de una ecuación de orden p .
- Afortunadamente es muy simple expresarla como una ecuación vectorial en diferencia de orden 1, que se resuelve de manera similar a la ecuación escalar.

Solución de la ecuación de orden p

Para resolverla se definen:

$$\xi_t \equiv \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad F \equiv \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad v_t \equiv \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

con lo que la ecuación de orden p que puede escribirse:

$$\xi_t = F\xi_{t-1} + v_t$$

y resolverse como:

$$\xi_{t+j} = F^{j+1}\xi_{t-1} + F^j v_t + F^{j-1} v_{t+1} + \cdots + F v_{t+j-1} + v_{t+j}$$

Nota: Descomposición espectral de una matriz

Descomposición espectral de una matriz

Si los eigenvectores de la matriz cuadrada A son linealmente independientes, entonces

$$A = C\Lambda C^{-1}$$

donde Λ es la matriz diagonal formada por los eigenvalores de A :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

y las columnas de C son los correspondientes eigenvectores de A .

Potencia de una matriz

Si A tiene la descomposición espectral $A = C\Lambda C^{-1}$ es fácil calcular su t -ésima potencia:

$$A^t = C\Lambda^t C^{-1}$$

ya que

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & 0 & \dots & 0 \\ 0 & \lambda_2^t & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n^t \end{bmatrix}$$

Ejemplo 2-1: Resolviendo $y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3$

La ecuación puede escribirse como

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \xi_t \end{bmatrix} = \begin{bmatrix} 0.9 & -0.2 \\ 1 & 0 \\ F & \xi_{t-1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ v \end{bmatrix}$$

Por sustitución recursiva encontramos

$$\begin{aligned} \xi_t &= (I + F + F^2 + \dots + F^{t-2})v + F^{t-1}\xi_1 \\ &= (I - F)^{-1}(I - F)(I + F + F^2 + \dots + F^{t-2})v + F^{t-1}\xi_1 \\ &= (I - F)^{-1}(I - F^{t-1})v + F^{t-1}\xi_1 \end{aligned}$$

Necesitamos una expresión para F^{t-1} . Para ello, buscamos la descomposición espectral de F .

Encontramos los eigenvalores de F :

$$\begin{vmatrix} 0.9 - \lambda & -0.2 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 0.9\lambda + 0.2 = (\lambda - 0.5)(\lambda - 0.4) = 0$$

es decir, $\lambda_1 = 0.5, \lambda_2 = 0.4$.

Es fácil mostrar que los eigenvectores son $\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ y $\begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$.

Entonces

$$\begin{aligned} F &= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \Rightarrow \\ F^{t-1} &= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{t-1} & 0 \\ 0 & \lambda_2^{t-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \Rightarrow \\ &= \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{t-1} & 0 \\ 0 & \lambda_2^{t-1} \end{bmatrix} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix} \\ &= \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} \lambda_1^t - \lambda_2^t & -\lambda_1 \lambda_2 (\lambda_1^{t-1} - \lambda_2^{t-1}) \\ \lambda_1^{t-1} - \lambda_2^{t-1} & -\lambda_1 \lambda_2 (\lambda_1^{t-2} - \lambda_2^{t-2}) \end{bmatrix} \end{aligned}$$

Sustituyendo los valores de λ_1 y λ_2

$$F^{t-1} = \begin{bmatrix} 10(0.5^t - 0.4^t) & -2(0.5^{t-1} - 0.4^{t-1}) \\ 10(0.5^{t-1} - 0.4^{t-1}) & -2(0.5^{t-2} - 0.4^{t-2}) \end{bmatrix}$$

Por otra parte:

$$(I - F)^{-1} = \frac{10}{3} \begin{bmatrix} 1 & -0.2 \\ 1 & 0.1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 10 & -2 \\ 10 & 1 \end{bmatrix}$$

Sustituyendo lo anterior en la ecuación vectorial:

$$\begin{aligned} \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} &= \frac{1}{3} \begin{bmatrix} 10 & -2 \\ 10 & 1 \end{bmatrix} \begin{bmatrix} 1-10(0.5^t-0.4^t) & 2(0.5^{t-1}-0.4^{t-1}) \\ -10(0.5^{t-1}-0.4^{t-1}) & 1-2(0.5^{t-2}-0.4^{t-2}) \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \\ &\quad \begin{bmatrix} 10(0.5^t-0.4^t) & -2(0.5^{t-1}-0.4^{t-1}) \\ 10(0.5^{t-1}-0.4^{t-1}) & -2(0.5^{t-2}-0.4^{t-2}) \end{bmatrix} \begin{bmatrix} y_1 \\ y_0 \end{bmatrix} = \\ &\quad \begin{bmatrix} 10 & -2 \\ 10 & 1 \end{bmatrix} \begin{bmatrix} 1-10(0.5^t-0.4^t) \\ -10(0.5^{t-1}-0.4^{t-1}) \end{bmatrix} + \begin{bmatrix} 10(0.5^t-0.4^t) & -2(0.5^{t-1}-0.4^{t-1}) \\ 10(0.5^{t-1}-0.4^{t-1}) & -2(0.5^{t-2}-0.4^{t-2}) \end{bmatrix} \begin{bmatrix} y_1 \\ y_0 \end{bmatrix} \end{aligned}$$

Tomando la primera fila

$$y_t = 10 - 100(0.5^t - 0.4^t) + 20(0.5^{t-1} - 0.4^{t-1}) + 10y_1(0.5^t - 0.4^t) - 2y_0(0.5^{t-1} - 0.4^{t-1}) \Rightarrow$$

$$\begin{aligned} y_t &= 10 + 10(y_1 - 10)(0.5^t - 0.4^t) - 2(y_0 - 10)(0.5^{t-1} - 0.4^{t-1}) \\ &= 10 + (10y_1 - 100)(0.5^t - 0.4^t) + (20 - 2y_0)(2 \times 0.5^t - 2.5 \times 0.4^t) \\ &= 10 + (10y_1 - 4y_0 - 60)0.5^t - (10y_1 - 5y_0 - 50)0.4^t \end{aligned}$$

Si imponemos las 2 condiciones iniciales: $y_0 = 13, y_1 = 11.3$, la solución de la ecuación es:

$$y_t = 10 + 0.5^t + 2 \times 0.4^t$$

Multiplicador dinámico: caso orden p

- De nuevo el multiplicador dinámico se obtiene por derivación:

$$\frac{\partial \xi_{t+j}}{\partial v'_t} = F^j$$

- El primer elemento de ξ_{t+j} es y_{t+j} y el primer elemento de v_t es w_t , por lo tanto:

$$\frac{\partial y_{t+j}}{\partial w_t} = F_{(11)}^j$$

- Ahora la estabilidad depende de F^j .

Estabilidad

- Que F^j tienda a 0 cuando j crece al infinito depende de los eigenvalores de F .
- Si todos son distintos, entonces $F^j = T\Lambda^j T^{-1}$, donde:

$$F^j = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1p} \\ t_{21} & t_{22} & \dots & t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{p1} & t_{p2} & \dots & t_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1^j & 0 & \dots & 0 \\ 0 & \lambda_2^j & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p^j \end{bmatrix} \begin{bmatrix} t^{11} & t^{12} & \dots & t^{1p} \\ t^{21} & t^{22} & \dots & t^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t^{p1} & t^{p2} & \dots & t^{pp} \end{bmatrix}$$

- por tanto

$$\begin{aligned} \frac{\partial y_{t+j}}{\partial w_t} &= F_{(11)}^j = (t_{11}t^{11})\lambda_1^j + (t_{12}t^{21})\lambda_2^j + \dots + (t_{1p}t^{p1})\lambda_p^j \\ &= c_1\lambda_1^j + c_2\lambda_2^j + \dots + c_p\lambda_p^j \end{aligned}$$

Obteniendo los eigenvalores y los ponderadores

- Los eigenvalores de F se obtienen de resolver:

Ecuación característica

$$|F - \lambda I_p| = \lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \dots - \phi_{p-1}\lambda - \phi_p = 0$$

- Mientras que c_i se obtiene de:

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{\substack{k=1 \\ k \neq i}}^p (\lambda_i - \lambda_k)}$$

- Nótese que:

$$c_1 + c_2 + \dots + c_p = (t_{11}t^{11}) + (t_{12}t^{21}) + \dots + (t_{1p}t^{p1}) = 1$$

Dinámica de ajuste

- Dado que

$$\begin{aligned} \frac{\partial y_{t+j}}{\partial w_t} &= c_1\lambda_1^j + c_2\lambda_2^j + \dots + c_p\lambda_p^j \\ 1 &= c_1 + c_2 + \dots + c_p \end{aligned}$$

el multiplicador dinámico es un promedio ponderado de las potencias de los eigenvalores.

- La forma del ajuste dependerá del eigenvalor de mayor valor absoluto λ_{max}

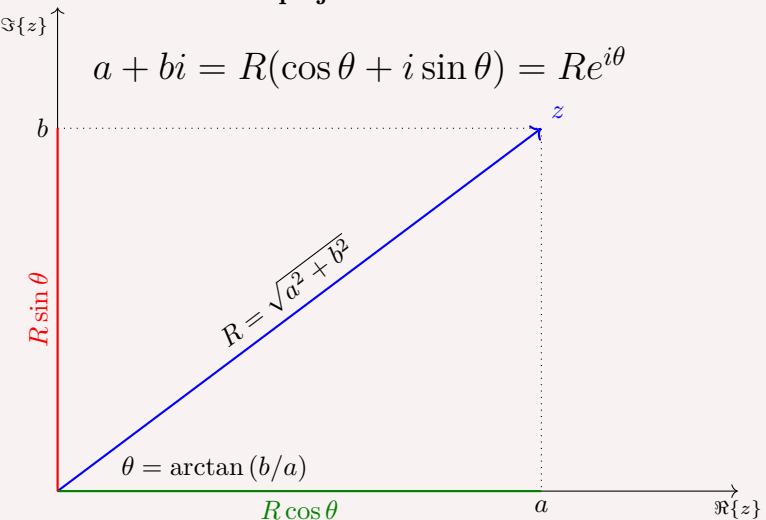
– Si $0 < \lambda_{max} < 1$, el MD decae geométricamente.

– Si $-1 < \lambda_{max} < 0$, el MD decae alternando

– Si $|\lambda_{max}| > 1$, la serie explota (no converge)

Nota: Números complejos

Representación de números complejos



Multiplicación de números complejos

- Si $z = Re^{i\theta}$ y $w = Se^{i\varphi}$, entonces su producto es

$$zw = RSe^{i(\theta+\varphi)}$$

- Así, si elevamos z a la n -ésima potencia:

$$z^n = (Re^{i\theta})^n = R^n e^{in\theta}$$

- Es decir

$$\lim_{n \rightarrow \infty} z^n = 0 \Leftrightarrow |R| < 1$$

Figura 2.1: Ejemplos de potencia de números complejos

Dinámica de ajuste (2)

- ¿Cómo se da el ajuste si λ_{max} es complejo?
- Se sabe que si $\lambda_1 = a + bi$, entonces $\lambda_2 = a - bi$
- Si expresamos λ_j en coordenadas polares:

$$\begin{aligned}\lambda_1 &= R[\cos \theta + i \sin \theta] = Re^{i\theta} \\ \lambda_2 &= R[\cos \theta - i \sin \theta] = Re^{-i\theta}\end{aligned}$$

$$\begin{aligned}R &= \sqrt{a^2 + b^2} \\ a &= R \cos \theta \\ b &= R \sin \theta\end{aligned}$$

- De lo anterior:

$$\begin{aligned}\lambda_1^j &= R^j e^{ij\theta} = R^j [\cos j\theta + i \sin j\theta] \\ \lambda_2^j &= R^j e^{-ij\theta} = R^j [\cos j\theta - i \sin j\theta]\end{aligned}$$

Dinámica de ajuste (3)

- El promedio de estos dos eigenvalores es:

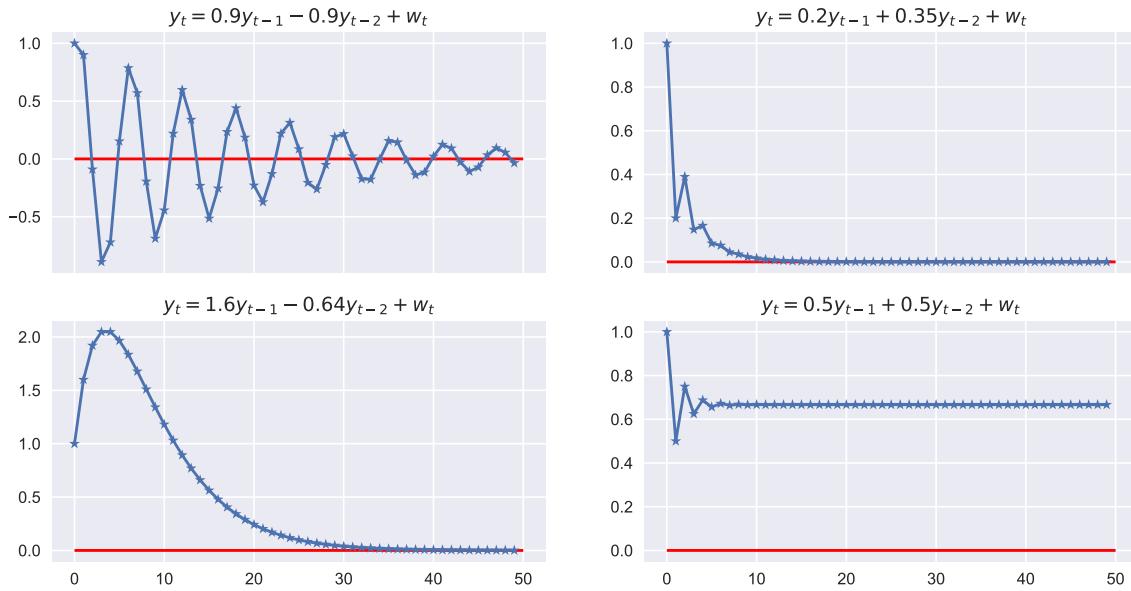
$$\begin{aligned}c_1 \lambda_1^j + c_2 \lambda_2^j &= c_1 R^j [\cos j\theta + i \sin j\theta] + c_2 R^j [\cos j\theta - i \sin j\theta] \\ &= R^j [(c_1 + c_2) \cos j\theta + i(c_1 - c_2) \sin j\theta]\end{aligned}$$

Pero c_1 y c_2 son conjugados: $c_1, c_2 = \alpha \pm \beta i$

$$= R^j [2\alpha \cos j\theta - 2\beta \sin j\theta]$$

- que en función de j es periódica, con frecuencia θ y período $\frac{2\pi}{\theta}$

Ejemplo de dinámica de ajuste cuando $p = 2$



Valor presente

- Recordando que

$$\frac{\partial \xi_{t+j}}{\partial v'_t} = F^j$$

Se tiene que:

$$\sum_{j=0}^{\infty} \beta^j \frac{\partial \xi_{t+j}}{\partial v'_t} = \sum_{j=0}^{\infty} \beta^j F^j = (I_p - \beta F)^{-1}$$

En este caso su elemento 1,1 es

$$\frac{1}{1 - \phi_1\beta - \phi_2\beta^2 - \cdots - \phi_p\beta^p}$$

Efecto acumulado y multiplicador de largo plazo

- Se obtiene del VP en el caso particular en que $\beta = 1$:

$$\frac{1}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

2.2 Solución por combinación de soluciones homogéneas y particulares

La estrategia de solución

Para resolver la ecuación lineal en diferencia

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

seguimos estos pasos

Paso 1: Formamos la ecuación homogénea $y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = 0$ y encontramos sus p soluciones;

Paso 2: Encontramos una solución particular;

Paso 3: Obtenemos la solución general como la suma de la solución particular y una combinación lineal de todas las soluciones homogéneas;

Paso 4: Eliminamos las constantes arbitrarias imponiendo p condiciones iniciales en el problema.

Ecuación homogénea de primer orden

- Para la ecuación

$$y_t = \phi y_{t-1} \Rightarrow y_t - \phi y_{t-1} = 0$$

- Solución trivial: $y_t = y_{t-1} = \dots = 0$, pero no es única.

- La expresión $y_t^h = \phi^t$ también es una solución:

$$\begin{array}{c} \phi^t - \phi(\phi^{t-1}) \\ \textcolor{teal}{y_t^h} \qquad \qquad \textcolor{teal}{y_{t-1}^h} \end{array} = 0$$

- Pero si y_t^h es una solución, entonces Ay_t^h también lo es, para cualquier escalar A :

$$Ay_t^h - \phi(Ay_{t-1}^h) = A(y_t^h - \phi y_{t-1}^h) = 0$$

Condición inicial para la ecuación homogénea de primer orden

- Hemos obtenido que $y_t = A\phi^t$ resuelve $y_t - \phi y_{t-1} = 0$
- Para determinar un valor específico de A , necesitamos una condición inicial.
- Por ejemplo, supongamos que el valor de y_t en $t = 0$ es conocido. Entonces:

$$y_0 = A\phi^0 \Rightarrow A = y_0$$

- Por lo que en ese caso la solución de la ecuación sería

$$y_t = \phi^t y_0$$

Ecuación homogénea de orden p

- Para la ecuación

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_{p-1} y_{t-p+1} - \phi_p y_{t-p} = 0$$

- Solución trivial es de nuevo: $y_t = y_{t-1} = \dots = 0$.

- Supongamos que la expresión $y_t^h = z^t$ también es una solución. Sustituyendo en la ecuación:

$$\begin{aligned} z^t - \phi_1 z^{t-1} - \phi_2 z^{t-2} - \dots - \phi_{p-1} z^{t-p+1} - \phi_p z^{t-p} &= 0 \\ z^{t-p} [z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_{p-1} z^1 - \phi_p z^0] &= 0 \end{aligned}$$

- Hemos logrado cambiar el problema original por el de encontrar los ceros de un polinomio de grado p :

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \dots - \phi_{p-1} z - \phi_p = 0$$

- Esta es la misma ecuación característica que encontramos en la sección anterior.

Resolviendo la ecuación característica

- Todo polinomio de grado p tiene exactamente p raíces, no necesariamente distintas o reales.
- Supongamos que z_1, z_2, \dots, z_p son las raíces del polinomio.
- Las soluciones homogéneas son entonces

$$y_t^h \in \{z_1^t, z_2^t, \dots, z_p^t\}$$

- Cualquier combinación lineal de estas soluciones $y_t^h = A_1 z_1^t + \dots + A_p z_p^t$ también es una solución:

$$\begin{aligned} y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} &= (A_1 z_1^t + \dots + A_p z_p^t) - \\ &\quad \phi_1 (A_1 z_1^{t-1} + \dots + A_p z_p^{t-1}) - \dots - \phi_p (A_1 z_1^{t-p} + \dots + A_p z_p^{t-p}) = \\ &= A_1 (z_1^t - \phi_1 z_1^{t-1} - \dots - \phi_p z_1^{t-p}) + \dots + A_p (z_p^t - \phi_1 z_p^{t-1} - \dots - \phi_p z_p^{t-p}) = \\ &= A_1 z_1^{t-p} (z_1^p - \phi_1 z_1^{p-1} - \dots - \phi_p) + \dots + A_p z_p^{t-p} (z_p^p - \phi_1 z_p^{p-1} - \dots - \phi_p) = 0 \end{aligned}$$

Ejemplo 2-2: Resolviendo $y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3$

 EqDiff2.ipynb

$$y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3$$

Paso 1: Resolvemos la ecuación homogénea $y_t - 0.9y_{t-1} + 0.2y_{t-2} = 0$:

$$\begin{aligned} z^2 - 0.9z + 0.2 &= (z - 0.4)(z - 0.5) = 0 \\ z \in \{0.4, 0.5\} &\Rightarrow y_{1,t}^h = 0.4^t, y_{2,t}^h = 0.5^t \end{aligned}$$

Es fácil verificar que son las soluciones:

$$\begin{aligned} 0.4^t - 0.9(0.4)^{t-1} + 0.2(0.4)^{t-2} &= (0.4)^{t-2} [(0.4)^2 - 0.9(0.4) + 0.2] = 0 \\ 0.5^t - 0.9(0.5)^{t-1} + 0.2(0.5)^{t-2} &= (0.5)^{t-2} [(0.5)^2 - 0.9(0.5) + 0.2] = 0 \end{aligned}$$

Paso 2: Supongamos que $y_t^p = c$, una constante, es una solución particular:

$$c = 0.9c - 0.2c + 3 \Rightarrow c = 10 \Rightarrow y_t^p = 10$$

Paso 3: Obtenemos la solución general como la suma de la solución particular y una combinación lineal de todas las soluciones homogéneas:

$$y_t = A_1(0.4)^t + A_2(0.5)^t + 10$$

Paso 4: Eliminamos A_1, A_2 imponiendo 2 condiciones iniciales: $y_0 = 13, y_1 = 11.3$.

$$\begin{cases} 13 &= A_1 + A_2 + 10 \\ 11.3 &= 0.4A_1 + 0.5A_2 + 10 \end{cases} \Rightarrow \begin{cases} A_1 + A_2 &= 3 \\ 0.4A_1 + 0.5A_2 &= 1.3 \end{cases}$$

entonces $A_1 = 2, A_2 = 1$ y la solución general de la ecuación es:

$$y_t = 2(0.4)^t + (0.5)^t + 10$$

Podemos también resolver este sistema utilizando el paquete `sympy` de Python:

```

1 from sympy import Function, rsolve
2 from sympy.abc import t
3
4 y = Function('y')
5 rsolve(y(t) - 0.9*y(t-1) + 0.2*y(t-2) - 3, y(t), {y(0):13, y(1):
    11.3})

```

2.3 Solución por medio del operador de rezagos

Introducción

- Este tema constituye una herramienta para simplificar el análisis de ecuaciones en diferencia
- Se asumirá que todas las variables son determinísticas (no estocásticas).
- Se limita la exposición a ecuaciones en diferencia lineales.

Ecuación en diferencia de primer orden

- En este caso

$$y_t = \phi y_{t-1} + w_t$$

- Utilizando el operador de rezagos se resuelve así:

$$\begin{aligned} y_t &= \phi L y_t + w_t \\ (1 - \phi L) y_t &= w_t \\ (1 + \phi L + \dots + \phi^t L^t) (1 - \phi L) y_t &= (1 + \phi L + \dots + \phi^t L^t) w_t \\ (1 - \phi^{t+1} L^{t+1}) y_t &= w_t + \phi w_{t-1} + \dots + \phi^t w_0 \end{aligned}$$

- Así

$$y_t = \phi^{t+1} y_{-1} + w_t + \phi w_{t-1} + \dots + \phi^t w_0$$

Solución de “largo plazo”

- En este caso

$$\begin{aligned} (1 - \phi L) y_t &= w_t \\ (1 - \phi L)^{-1} (1 - \phi L) y_t &= (1 + \phi L + \phi^2 L^2 + \dots) w_t \\ y_t &= w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \dots \end{aligned}$$

siempre y cuando $|\phi| < 1$.

Ecuación en diferencia de orden p

- La variable y_t evoluciona como

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

- Con operador de rezagos:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) y_t = w_t$$

- Para factorizar el polinomio es necesario resolver

$$f(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

- Con el cambio de variable $z = \frac{1}{\lambda}$ obtenemos

$$\begin{aligned} 1 - \phi_1 \left(\frac{1}{\lambda}\right) - \phi_2 \left(\frac{1}{\lambda}\right)^2 - \cdots - \phi_p \left(\frac{1}{\lambda}\right)^p &= 0 \\ \lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_{p-1} \lambda - \phi_p &= 0 \end{aligned}$$

- Esta es la misma expresión que se obtuvo con álgebra de matrices: por lo tanto las raíces de $f(z)$ son los *recíprocos* de las raíces anteriores.

Estabilidad

- Dada la relación existente entre las ecuaciones

$$\begin{aligned} 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p &= 0 \\ \lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_{p-1} \lambda - \phi_p &= 0 \end{aligned}$$

está claro que para que el proceso sea estable es necesario que las raíces de

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

estén *fuera* del círculo unitario, esto es, si z_i es raíz, entonces $|z_i| > 1$.

Referencias del capítulo 2

Enders, Walter (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
 Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.

Capítulo 3

Modelos AutoRegresivos de Media Móvil (ARMA)

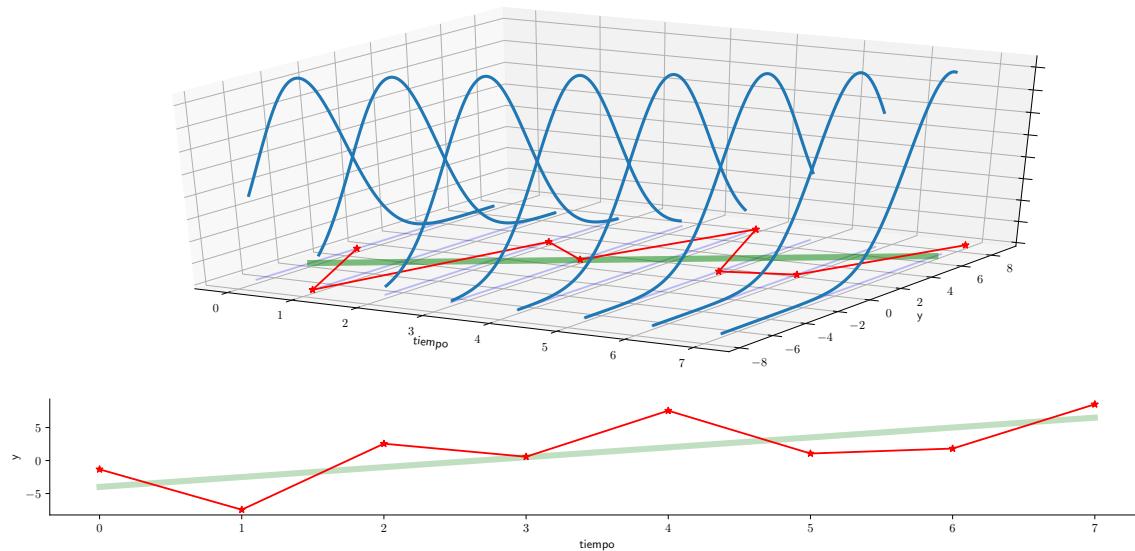
En este capítulo

3.1	Introducción	39
3.2	Proceso de media móvil MA(q)	44
3.3	Proceso autorregresivo AR(p)	49
3.4	Proceso autorregresivo de media móvil ARMA(p,q)	57
3.5	Estimación de modelos ARMA	59
3.6	Pronósticos con modelos ARMA	65

3.1 Introducción

¿Qué es una serie de tiempo?

Una serie de tiempo $\{y_t\}_{t=1}^T$ es una realización de un proceso estocástico.



¿Qué nos gustaría hacer con esta serie de tiempo?

- Imaginemos a la serie de tiempo como la parte de un proceso estocástico para la cual ya tenemos las realizaciones del proceso (un valor por período)

$$\underbrace{y_1, y_2, \dots, y_T}_{\text{datos observados}}, \underbrace{y_{T+1}, y_{T+2}, \dots}_{\text{futuro}}$$

- Nos gustaría saber *la distribución condicional*

$$\mathbb{P}[y_{T+j} \leq x | y_1, \dots, y_T]$$

- Esto nos permitiría utilizar nuestra serie de tiempo para pronosticar valores futuros de la serie, así como precisar su variabilidad:

$$\begin{aligned}\mathbb{E}_T[y_{T+j}] &\equiv \mathbb{E}[y_{T+j} | y_1, \dots, y_T] \\ \text{Var}_T[y_{T+j}] &\equiv \text{Var}[y_{T+j} | y_1, \dots, y_T]\end{aligned}$$

Nuestra tarea

- En la práctica, puede ser que nunca conozcamos el verdadero *proceso generador de datos (PGD)* (el proceso estocástico del cual fueron obtenidos los valores de nuestra serie de tiempo).
- Nuestra tarea es desarrollar *modelos* que capturen la esencia del verdadero PGD.
- Las ecuaciones en diferencia estocásticas son una manera muy conveniente de modelar procesos económicos dinámicos.

Ejemplo 3-1: Controlando la oferta de dinero

- Suponga que la meta de oferta monetaria M^* del banco central crece 100g% por año:

$$M_t^* = (1 + g)M_{t-1}^*$$

o en términos logarítmicos, con $m^* \equiv \log(M^*)$

$$\begin{aligned}m_t^* &= \log(1 + g) + m_{t-1}^* \\ m_t^* &\approx g + m_{t-1}^*\end{aligned}$$

- Para una condición inicial m_0^* dada, la solución es:

$$m_t^* = gt + m_0^*$$

- La cantidad efectiva de dinero m_t puede diferir de la meta m_t^* .
- El banco central intenta cerrar una proporción ρ de la brecha entre la meta y la cantidad efectiva del período anterior. El cambio en la oferta de dinero es:

$$\Delta m_t = \rho(m_{t-1}^* - m_{t-1}) + \epsilon_t$$

política monetaria perturbación

- por lo que la oferta de dinero es

$$\begin{aligned}
 m_t &= \rho m_{t-1}^* + (1 - \rho)m_{t-1} + \epsilon_t \\
 &= \rho g(t-1) + \rho m_0^* + (1 - \rho)m_{t-1} + \epsilon_t \\
 &= \underbrace{\rho(m_0^* - g)}_{\text{intercepto}} + \underbrace{(\rho g)t}_{\text{tendencia}} + \underbrace{(1 - \rho)m_{t-1}}_{\text{autorregresivo}} + \underbrace{\epsilon_t}_{\text{shock}}
 \end{aligned}$$

- Aunque la oferta monetaria es una variable continua, nuestro modelo es una ecuación en diferencia (discreta).
- Como las perturbaciones $\{\epsilon_t\}$ son aleatorias, la oferta de dinero es estocástica.
- Si supiéramos la distribución de $\{\epsilon_t\}$, podríamos calcular la distribución de cada elemento de $\{m_t\}$, porque está determinada completamente por los parámetros de la ecuación y por la secuencia $\{\epsilon_t\}$.
- Habiendo observado las primeras T observaciones de la serie $\{m_t\}$, podríamos pronosticar futuros valores. Por ejemplo:

$$\begin{aligned}
 m_{T+1} &= \rho(gT + m_0^*) + (1 - \rho)m_T + \epsilon_{T+1} \\
 \Rightarrow \mathbb{E}_T[m_{T+1}] &= \rho(gT + m_0^*) + (1 - \rho)m_T
 \end{aligned}$$

Ruido blanco y el modelo clásico de regresión lineal

- En el MCRL se asume que

$$y_t = x_t' \beta + \epsilon_t$$

donde para todas las observaciones $t = 1, 2, \dots, T$ el término de error cumple que:

$$\begin{aligned}
 \mathbb{E}[\epsilon_t] &= 0 && (\text{media cero}) \\
 \text{Var}[\epsilon_t] &= \sigma^2 && (\text{no hay heteroscedasticidad}) \\
 \mathbb{E}[\epsilon_t \epsilon_\tau] &= 0 \quad \text{si } t \neq \tau && (\text{no hay autocorrelación})
 \end{aligned}$$

- Es decir, el MCRL aplicado a series de tiempo asume que el error es un proceso de ruido blanco.
- Sin embargo, en la práctica rara vez se satisface ese supuesto cuando se ajusta un modelo de regresión lineal a datos de series de tiempo.

Ejemplo 3-2: Estimando la demanda de dinero

 bccr

 demandadinerо.ipynb

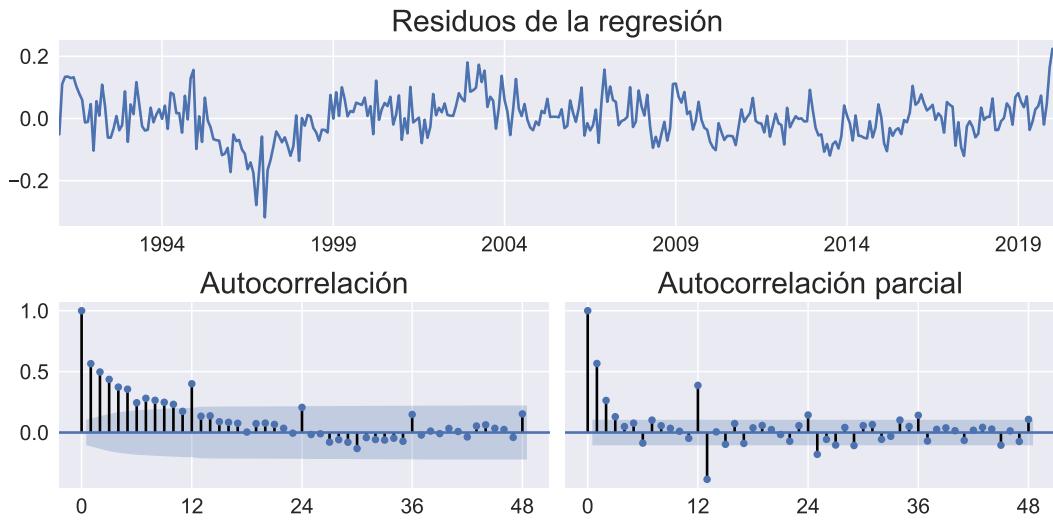
Consideremos la demanda de dinero en Costa Rica

$$\log(M_t) = \beta_0 + \beta_1 \log(q_t) + \beta_2 \log(p_t) + \beta_3 \log(i_t) + \epsilon_t$$

la cual estimamos con datos mensuales (1991m01 a 2020m01) del medio circulante M_t , el IMAE q_t , el IPC p_t , y la tasa básica pasiva i_t .

	coef	std err	t	P> t	[0.025	0.975]
Intercept	5.5080	0.227	24.251	0.000	5.061	5.955
IMAE	1.1300	0.059	19.143	0.000	1.014	1.246
IPC	0.9463	0.021	45.664	0.000	0.905	0.987
Tbasica	-0.2409	0.018	-13.641	0.000	-0.276	-0.206

No obstante, los residuos de la regresión no parecen ruido blanco.



- Conocer el valor de un residuo puede ayudar a pronosticar el siguiente.
- Aún así, notemos que para pronosticar el valor del M1 en 2020m02, necesitaríamos pronosticar los valores de las demás variables.

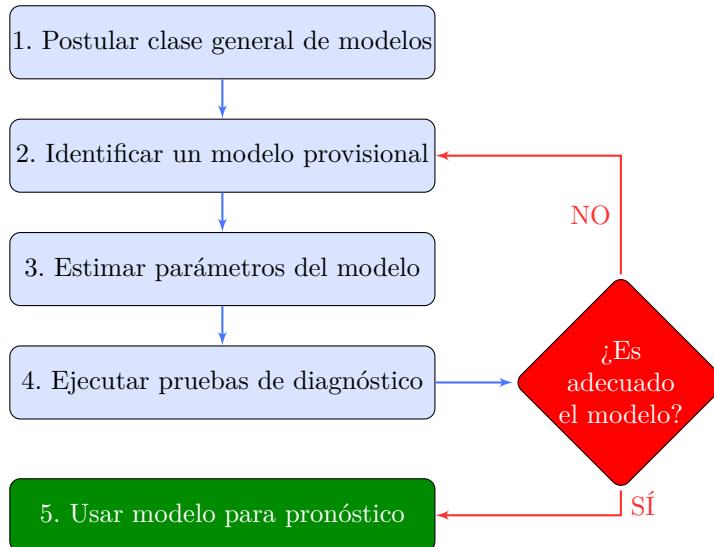
En esta clase

- En esta clase aprenderemos a modelar series de tiempo en función de:
 - sus valores rezagados (procesos autorregresivos)
 - valores rezagados de un ruido blanco (procesos de media móvil)
- Primero estudiamos las propiedades teóricas de procesos estocásticos.
- Luego tratamos de identificar el PGD de nuestra serie a partir de sus estadísticos muestrales, comparándolos con los estadísticos de los procesos del punto anterior.
- Finalmente, utilizamos nuestro modelo estimado para
 - análisis de escenarios: ¿qué pasaría con la serie de tiempo si recibe una perturbación estocástica de cierta magnitud (función impulso respuesta)
 - pronósticos: ¿qué valores esperamos ver en el futuro para esta serie de tiempo?

Modelos que estudiaremos

<p>Ruido blanco</p> <p>Es una secuencia $\{\epsilon_t\}$ cuyos elementos satisfacen,</p> $\mathbb{E}(\epsilon_t) = 0$ $\mathbb{E}(\epsilon_t^2) = \sigma^2$ $\mathbb{E}(\epsilon_t \epsilon_\tau) = 0 \quad \text{for } t \neq \tau$	<p>Proceso media móvil</p> <p>Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico</p> $y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$ <p>con $\theta_q \neq 0$ es llamado un proceso MA(q).</p>
<p>Proceso autorregresivo</p> <p>Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico</p> $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$ <p>con $\phi_p \neq 0$ es llamado un proceso AR(p).</p>	<p>Autorregresivo media móvil</p> <p>Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico</p> $y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \cdots$ $+ \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$ <p>es llamado proceso ARMA(p,q).</p>

La metodología Box-Jenkins

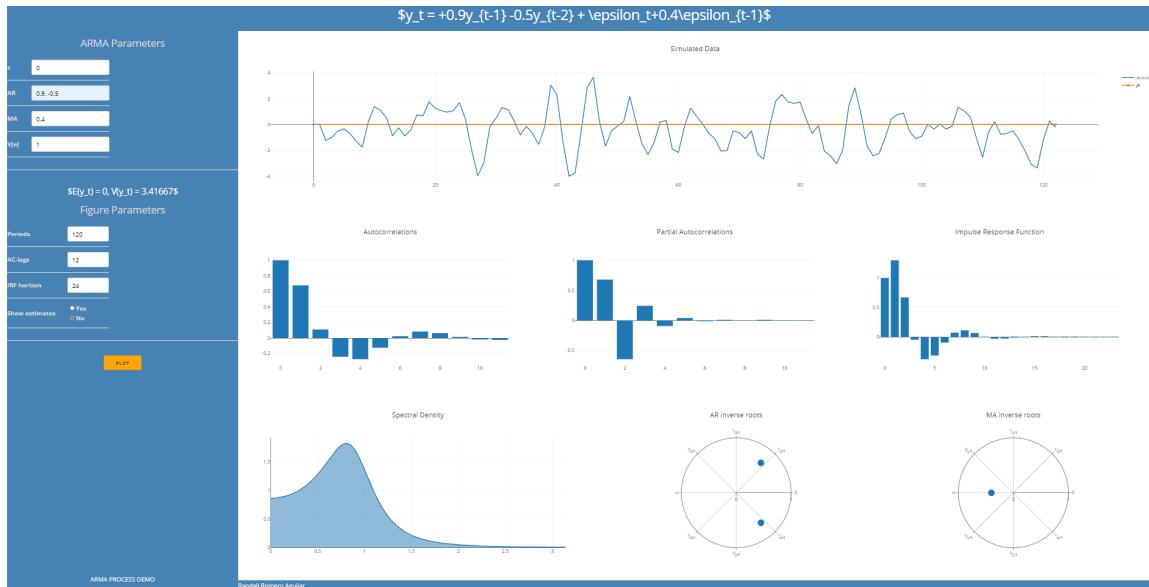


Acerca de los ejemplos

- En esta clase veremos ilustraciones de distintos proceso AR, MA, y ARMA.
- Usted puede reproducirlas (y estudiar más casos específicos de estos procesos) con el paquete **macrodemos** que escribí en Python para este tema.
- Para instalarlo: `pip install macrodemos`
- Para ejecutarlo:

```

1 from macrodemos import ARMA_demo
2 ARMA_demo()
  
```



3.2 Proceso de media móvil MA(q)

Proceso media móvil de primer orden: MA(1)

- Sea $\{\epsilon_t\}_{t=-\infty}^{\infty}$ un proceso ruido blanco.
- Se define el proceso MA(1) como:

$$\begin{aligned} y_t &= \mu + \epsilon_t + \theta\epsilon_{t-1} \\ &= \mu + (1 + \theta L)\epsilon_t \end{aligned}$$

- Su valor esperado es:

$$\begin{aligned} \mathbb{E}[y_t] &= \mu + \mathbb{E}[\epsilon_t] + \theta \mathbb{E}[\epsilon_{t-1}] \\ &= \mu + 0 + 0\theta \\ &= \mu \end{aligned}$$

- Su varianza es

$$\begin{aligned} \text{Var}[y_t] &\equiv \mathbb{E}[(y_t - \mathbb{E}[y_t])^2] \\ &= \mathbb{E}[(y_t - \mu)^2] \\ &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})^2] \\ &= \mathbb{E}[\epsilon_t^2 + 2\theta\epsilon_t\epsilon_{t-1} + \theta^2\epsilon_{t-1}^2] \\ &= \mathbb{E}[\epsilon_t^2] + 2\theta \mathbb{E}[\epsilon_t\epsilon_{t-1}] + \theta^2 \mathbb{E}[\epsilon_{t-1}^2] \\ &= \sigma^2 + 2\theta \times 0\theta^2\sigma^2 \\ &= (1 + \theta^2)\sigma^2 \end{aligned}$$

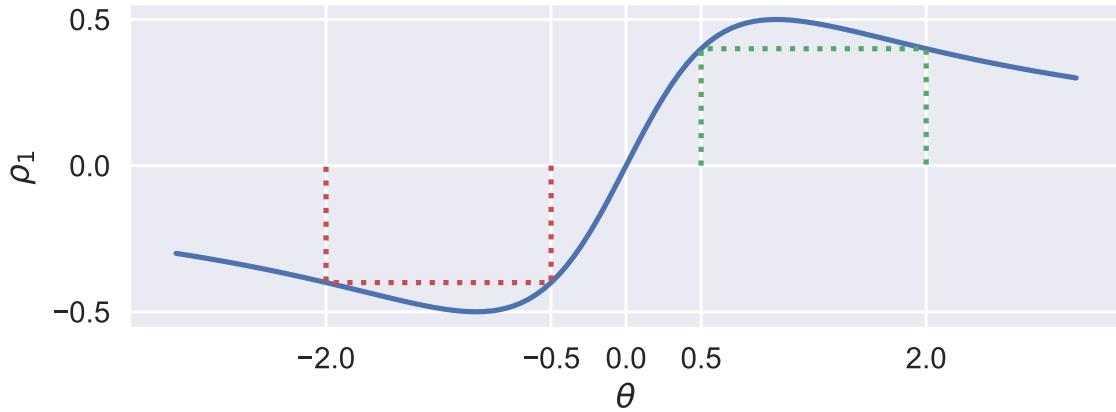
- Su autocovarianza j , para $j \geq 1$, es

$$\begin{aligned}
\text{Cov}[y_t, y_{t-j}] &= \mathbb{E}[(y_t - \mu)(y_{t-j} - \mu)] \\
&= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-j} + \theta\epsilon_{t-j-1})] \\
&= \mathbb{E}[\epsilon_t\epsilon_{t-j} + \theta\epsilon_t\epsilon_{t-j-1} + \theta\epsilon_{t-1}\epsilon_{t-j} + \theta^2\epsilon_{t-1}\epsilon_{t-j-1}] \\
&= 0 + \theta \times 0 + \theta \mathbb{E}[\epsilon_{t-1}\epsilon_{t-j}] + \theta^2 \times 0 \\
&= \theta \mathbb{E}[\epsilon_{t-1}\epsilon_{t-j}] = \begin{cases} \theta\sigma^2 & , \text{ si } j = 1 \\ 0 & , \text{ si } j > 1 \end{cases}
\end{aligned}$$

- Por lo tanto, su función de autocorrelación es

$$\rho_j = \begin{cases} \frac{\theta}{1+\theta^2} & , \text{ si } j = 1 \\ 0 & , \text{ si } j > 1 \end{cases}$$

- Notemos que la función de autocorrelación de $y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$ es la misma que para el proceso $z_t = \mu + \epsilon_t + \frac{1}{\theta}\epsilon_{t-1}$



- Resumiendo los resultados que hemos obtenido

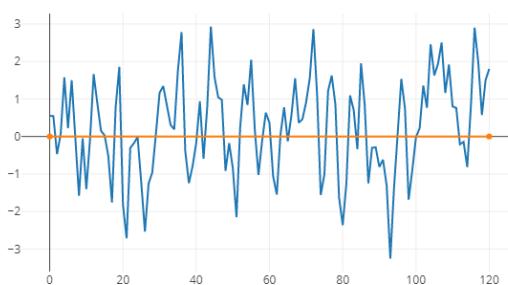
$$\begin{aligned}
\mathbb{E}[y_t] &= \mu \\
\text{Var}[y_t] &= (1 + \theta^2)\sigma^2 \\
\text{Cov}[y_t, y_{t-j}] &= \begin{cases} \theta\sigma^2 & , \text{ si } j = 1 \\ 0 & , \text{ si } j > 1 \end{cases}
\end{aligned}$$

vemos que ninguno de estos momentos depende del tiempo t , por lo que el proceso MA(1) siempre es covarianza-estacionario.

Ejemplo 3-3: MA(1)

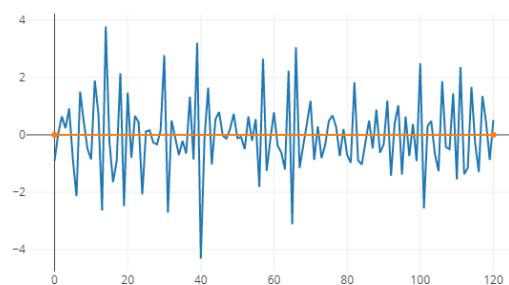
$$y_t = \epsilon_t + 0.8\epsilon_{t-1}$$

Realización

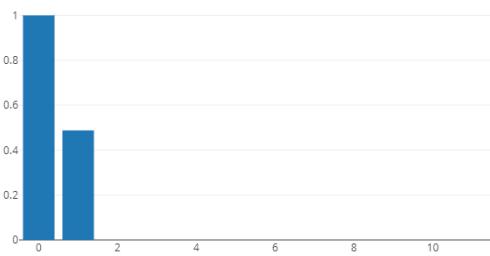


$$y_t = \epsilon_t - 0.8\epsilon_{t-1}$$

Realización



Autocorrelación



Autocorrelación



Invertibilidad de un proceso MA(1)

- Supongamos que $\mu = 0$, con lo que el proceso es

$$y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t$$

- Siempre que $|\theta| < 1$ podemos invertir el polinomio $(1 + \theta L)$:

$$(1 + \theta L)^{-1} = (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \dots)$$

- con lo que

$$\begin{aligned} \epsilon_t &= (1 + \theta L)^{-1} y_t \\ &= (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \dots) y_t \\ &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \dots \\ \Rightarrow y_t &= \epsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \dots \end{aligned}$$

- Es decir, podemos representar el proceso MA(1) como un proceso AR(∞).
- Recordemos que si bien un proceso MA(1) con parámetro θ tiene exactamente la misma función de autocorrelación que un proceso con parámetro $\frac{1}{\theta}$, solo uno de ellos puede ser invertible, porque si $|\theta| < 1$, entonces $|\frac{1}{\theta}| > 1$.
- Para ciertos métodos de estimación, sólo será posible estimar el modelo MA(1) si es invertible.
- Por ello, para modelos no invertibles se suele cambiar el parámetro por su recíproco.

El proceso MA(q)

- Es fácil extender el proceso MA(1) para incluir más rezagos.
- El proceso MA(q) es

$$\begin{aligned} y_t &= \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ &= \mu + (1 + \theta_1 L + \cdots + \theta_q L^q) \epsilon_t \\ &= \mu + \Theta(L) \epsilon_t \end{aligned}$$

con ϵ_t ruido blanco.

- Su valor esperado es

$$\begin{aligned} \mathbb{E}(y_t) &= \mathbb{E}(\mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}) \\ &= \mu + \mathbb{E}(\epsilon_t) + \theta_1 \mathbb{E}(\epsilon_{t-1}) + \cdots + \theta_q \mathbb{E}(\epsilon_{t-q}) \\ &= \mu \end{aligned}$$

- Su varianza es

$$\begin{aligned} \text{Var}[y_t] &= \mathbb{E}[(y_t - \mu)^2] \\ &= \mathbb{E}[(\epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q})^2] \\ &= \sigma^2 + \theta_1^2 \sigma^2 + \cdots + \theta_q^2 \sigma^2 \\ &= (1 + \theta_1^2 + \cdots + \theta_q^2) \sigma^2 \end{aligned}$$

- mientras que su autocovarianza es

$$\gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j}) \sigma^2 & \text{si } j = 1, 2, \dots, q \\ 0 & \text{para } j > q. \end{cases}$$

- es decir, una característica distintiva de un proceso MA(q) es que todas sus autocorrelaciones para rezagos mayores a q son cero.

Invertibilidad de un proceso MA(q)

- El proceso MA(q)

$$y_t = \mu + \Theta(L) \epsilon_t$$

será invertible si y solo si las raíces del polinomio $\Theta(z)$ están todas fuera del círculo unitario.

- En ese caso, el proceso se puede representar por

$$\epsilon_t = \Theta^{-1}(L)(y_t - \mu)$$

lo cual corresponde a un proceso AR(∞).

Función impulso respuesta de un proceso MA(q)

- La función de impulso respuesta está definida por

$$\Psi(j) = \frac{\partial y_{t+j}}{\partial \epsilon_t}$$

es decir, nos dice cuánto cambia y luego de j períodos ante una perturbación.

- Para series estacionarias, podemos escribir

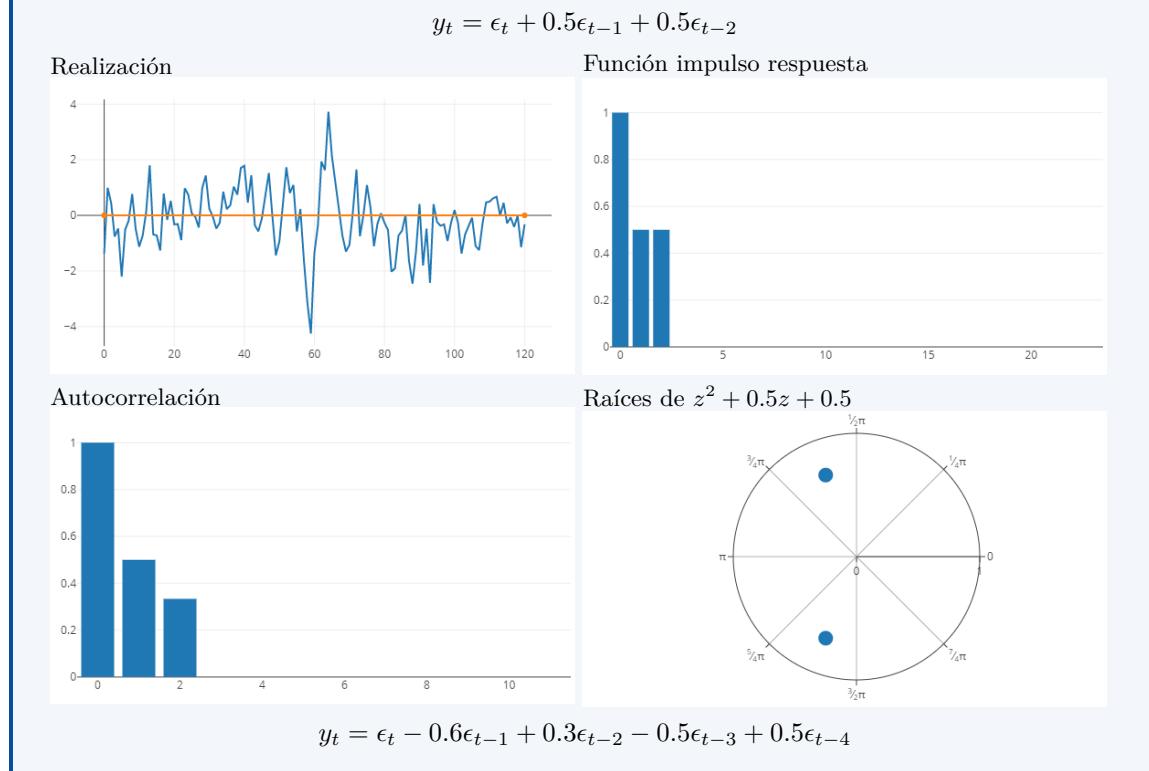
$$\Psi(j) = \frac{\partial y_t}{\partial \epsilon_{t-j}}$$

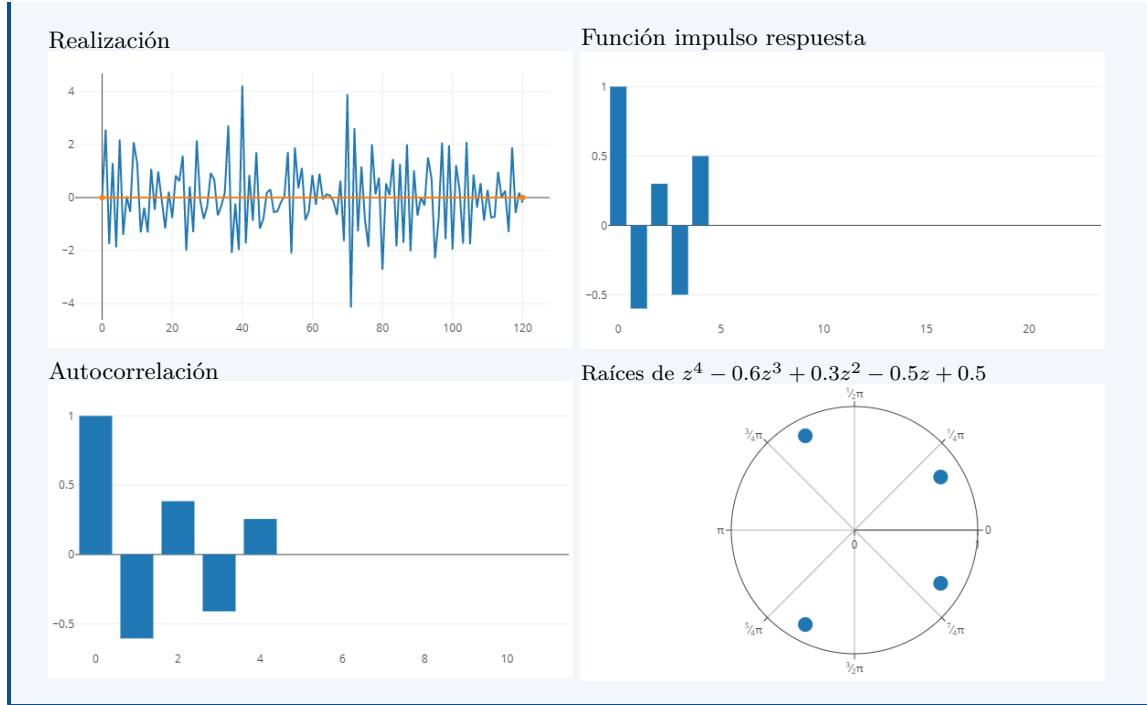
- Pero como $y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$ es fácil ver que

$$\Psi(j) = \theta_j$$

es decir, la función de impulso respuesta es idéntica a los coeficientes del proceso MA(q).

Ejemplo 3-4: MA(2) y MA(4)





Proceso media móvil de orden infinito: MA(∞)

- Definimos el proceso MA(∞) como

$$y_t = \mu + \psi_0 \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots$$

- Su media es

$$\mathbb{E}[y_t] = \mu$$

- Su varianza es

$$\gamma_0 = (\psi_0^2 + \psi_1^2 + \psi_2^2 + \dots) \sigma^2$$

la cual es finita siempre y cuando

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty$$

3.3 Proceso autorregresivo AR(p)

Proceso autorregresivo de primer orden: AR(1)

- Sea $\{\epsilon_t\}_{t=-\infty}^{\infty}$ un proceso ruido blanco.
- Se define el proceso AR(1) como:

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ y_t - \phi y_{t-1} &= c + \epsilon_t \\ (1 - \phi L)y_t &= c + \epsilon_t \end{aligned}$$

- Siempre que $|\phi| < 1$ podemos invertir el término $(1 - \phi L)$

$$\begin{aligned} y_t &= (1 - \phi L)^{-1} (c + \epsilon_t) \\ &= (1 - \phi L)^{-1} c + (1 - \phi L)^{-1} \epsilon_t \\ &= \frac{c}{1 - \phi} + (1 + \phi L + \phi^2 L^2 + \dots) \epsilon_t \\ &= \frac{c}{1 - \phi} + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots \end{aligned}$$

- Vemos por tanto que si $|\phi| < 1$ el proceso AR(1) puede escribirse como el proceso MA(∞)

$$y_t = \frac{c}{1 - \phi} + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots$$

- Por lo que su valor esperado es:

$$\mathbb{E}[y_t] \equiv \mu = \frac{c}{1 - \phi}$$

- y su varianza es

$$\begin{aligned} \text{Var}[y_t] &\equiv \gamma_0 = (1 + \phi^2 + \phi^4 + \phi^6 + \dots) \sigma^2 \\ &= \frac{\sigma^2}{1 - \phi^2} \end{aligned}$$

- La representación MA(∞) nos muestra que y_t depende de la perturbación presente ϵ_t y de *todas las pasadas* $\epsilon_{t-1}, \epsilon_{t-2}, \dots$, pero no depende de *ninguna de las perturbaciones futuras* $\epsilon_{t+1}, \epsilon_{t+2}, \dots$
- Por ello, la covarianza de y_t con cualquier perturbación posterior ϵ_{t+k} (con $k > 0$) es cero.
- Sabiendo que un proceso AR(1) es estacionario si $|\phi| < 1$, es fácil obtener sus momentos sin necesidad de obtener su representación MA(∞).
- En el caso de la media:

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ \mathbb{E}[y_t] &= c + \phi \mathbb{E}[y_{t-1}] + \mathbb{E}[\epsilon_t] \\ \mu &= c + \phi \mu + 0 \\ \Rightarrow \mu &= \frac{c}{1 - \phi} \end{aligned}$$

- Restando la tercera línea de la primera, vemos que el proceso puede escribirse como un AR(1) para la *desviación respecto a la media* sin la constante:

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t$$

- Para obtener su varianza elevamos al cuadrado la expresión anterior y tomamos su valor esperado

$$\begin{aligned} \mathbb{E}(\tilde{y}_t^2) &= \mathbb{E}[(\phi \tilde{y}_{t-1} + \epsilon_t)^2] \\ \gamma_0 &= \mathbb{E}[\phi^2 \tilde{y}_{t-1}^2 + 2\phi \tilde{y}_{t-1} \epsilon_t + \epsilon_t^2] \\ &= \phi^2 \gamma_0 + 2 \times 0 + \sigma^2 \\ \Rightarrow \gamma_0 &= \frac{\sigma^2}{1 - \phi^2} \end{aligned}$$

- Para obtener su autocovarianza γ_j , para $j \geq 1$, escribimos

$$\tilde{y}_t = \phi \tilde{y}_{t-1} + \epsilon_t$$

multiplicamos ambos lados por \tilde{y}_{t-j} y tomamos esperanza

$$\begin{aligned}\mathbb{E}[\tilde{y}_t \tilde{y}_{t-j}] &= \mathbb{E}[\phi \tilde{y}_{t-1} \tilde{y}_{t-j} + \tilde{y}_{t-j} \epsilon_t] \\ \gamma_j &= \phi \gamma_{j-1}\end{aligned}$$

- Vemos que la función de autocorrelación es la solución de una ecuación en diferencia homogénea de primer orden, cuya solución es

$$\gamma_j = \phi^j \gamma_0$$

- Resumiendo los resultados que hemos obtenido

$$\begin{aligned}\mathbb{E}[y_t] &= \mu = \frac{c}{1 - \phi} \\ \text{Var}[y_t] &= \gamma_0 = \frac{\sigma^2}{1 - \phi^2} \\ \text{Cov}[y_t, y_{t-j}] &= \gamma_j = \phi^j \gamma_0 \\ \rho_j &= \phi^j\end{aligned}$$

vemos que ninguno de estos momentos depende del tiempo t , por lo que el proceso AR(1) es covarianza-estacionario siempre y cuando $|\phi| < 1$.

- A partir de la representación MA(∞) del proceso AR(1) con $|\phi| < 1$

$$y_t = \mu + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots$$

es fácil observar que su función de impulso respuesta

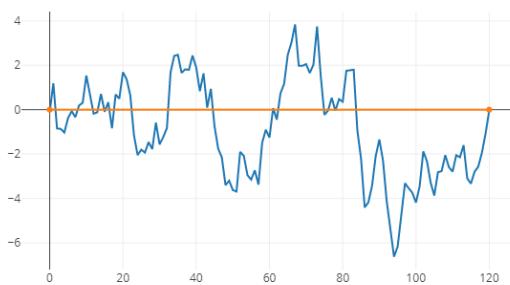
$$\Psi(j) = \frac{\partial y_{t+j}}{\partial \epsilon_t} = \frac{\partial y_t}{\partial \epsilon_{t-j}} = \phi^j$$

es idéntica a su función de autocorrelaciones.

Ejemplo 3-5: AR(1)

$$y_t = 0.9y_{t-1} + \epsilon_t$$

Realización

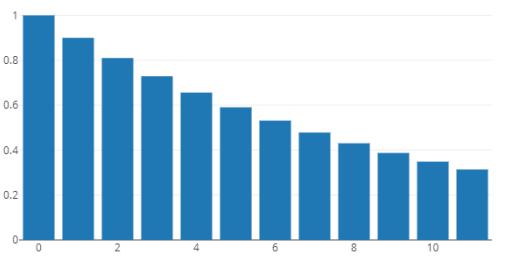


$$y_t = -0.9y_{t-1} + \epsilon_t$$

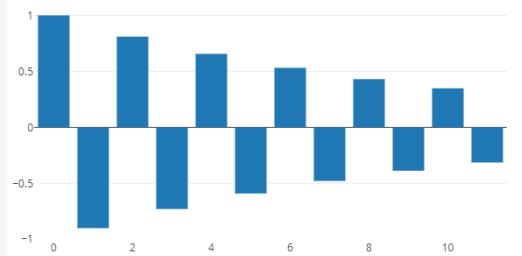
Realización



Autocorrelación



Autocorrelación



El proceso AR(p)

- Es fácil extender el proceso AR(1) para incluir más rezagos.
- El proceso AR(p) es

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \\ y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} &= c + \epsilon_t \\ (1 - \phi_1 L^1 - \cdots - \phi_p L^p) y_t &= c + \epsilon_t \\ \Phi(L) y_t &= c + \epsilon_t \end{aligned}$$

- El proceso AR(p) es una ecuación en diferencia de orden p .
- Esta ecuación es estable si y solo si las raíces del polinomio $1 - \phi_1 z^1 - \cdots - \phi_p z^p$ están todas fuera del círculo unitario.
- Si el proceso es estable, resolvemos para y_t

$$\begin{aligned} y_t &= \Phi^{-1}(L)(c + \epsilon_t) \\ &= \Phi^{-1}(1)c + \Phi^{-1}(L)\epsilon_t \\ &= \frac{c}{1 - \phi_1 - \cdots - \phi_p} + \Phi^{-1}(L)\epsilon_t \end{aligned}$$

- Su valor esperado es

$$\mathbb{E}(y_t) = \frac{c}{1 - \phi_1 - \cdots - \phi_p} = \mu$$

- Similar a lo que obtuvimos para el proceso AR(1), podemos escribir el proceso AR(p) en términos de desviación de la media $\tilde{y} \equiv y - \mu$.

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t$$

- Para obtener su varianza y autocovarianzas multiplicamos la expresión anterior por \tilde{y}_{t-j} , con $j \geq 0$, y calculamos el valor esperado

$$\begin{aligned}\mathbb{E}[\tilde{y}_t \tilde{y}_{t-j}] &= \mathbb{E}[\phi_1 \tilde{y}_{t-1} \tilde{y}_{t-j} + \cdots + \phi_p \tilde{y}_{t-p} \tilde{y}_{t-j} + \epsilon_t \tilde{y}_{t-j}] \\ \gamma_j &= \begin{cases} \phi_1 \gamma_1 + \cdots + \phi_p \gamma_p + \sigma^2 & \text{si } j = 0 \\ \phi_1 \gamma_{j-1} + \cdots + \phi_p \gamma_{j-p} & \text{si } j > 0 \end{cases}\end{aligned}$$

- Si dividimos la última línea entre la varianza γ_0 obtenemos las

Ecuaciones Yule-Walker

$$\rho_j = \phi_1 \rho_{j-1} + \cdots + \phi_p \rho_{j-p}, \text{ con } j > 0$$

- Esta es la misma ecuación en diferencia que el proceso original, por lo que en principio se puede resolver con los métodos convencionales, una vez que tengamos p valores iniciales $\rho_0, \rho_1, \dots, \rho_{p-1}$.
- En general, no es tan sencillo obtener despejar los valores de las autocorrelaciones ρ_j .
- Para obtener la función de impulso respuesta, podríamos partir de la forma

$$y_t = \mu + \Phi^{-1}(L)\epsilon$$

pero esto requiere que obtengamos la forma explícita del polinomio de rezagos

$$\Phi^{-1}(L) = 1 + \psi_1 L + \psi_2 L^2 + \dots$$

- Ahora bien, reconociendo que un proceso AR(p) es una ecuación en diferencia de orden p , encontramos que

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = c_1 \lambda_1^j + c_2 \lambda_2^j + \cdots + c_p \lambda_p^j$$

donde λ_j son las raíces del polinomio característico $\lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p$ (ver tema 2, p.19-20).

- Para un proceso AR(p) estacionario, sabemos que todas las raíces λ_j están dentro del círculo unitario.
- Por ello, a partir de

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = c_1 \lambda_1^j + c_2 \lambda_2^j + \cdots + c_p \lambda_p^j$$

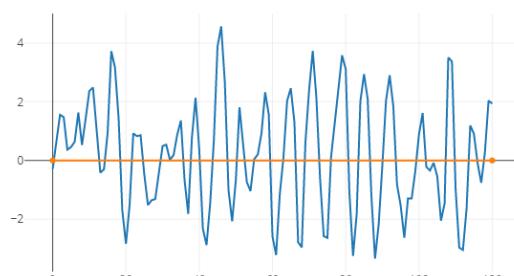
sabemos que la función de impulso respuesta converge a cero en el largo plazo.

- La forma de la función depende especialmente del valor λ_j más grande:
 - Si es positivo, decae geométricamente,
 - Si es negativo, decae alternando de signo,
 - Si es complejo, decae oscilando periódicamente (ver tema 2, p.25).

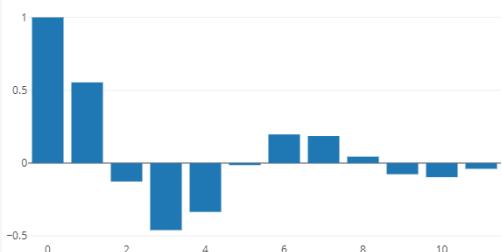
Ejemplo 3-6: AR(2): $y_t = 0.9y_{t-1} - 0.625y_{t-2} + \epsilon_t$

$$y_t = 0.9y_{t-1} - 0.625y_{t-2} + \epsilon_t$$

Realización



Autocorrelación



La ecuación característica es

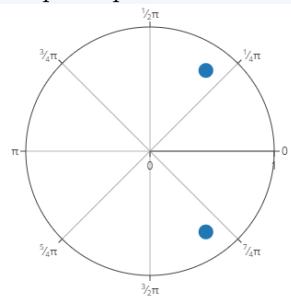
$$\lambda^2 - 0.9\lambda + 0.625 = 0$$

con raíces

$$\lambda = 0.45 \pm 0.65i$$

$$\approx 0.79 (\cos 0.97 \pm i \sin 0.97)$$

por lo que el proceso es estacionario.



Con las ecuaciones Yule-Walker

$$\phi_1 + \phi_2 \rho_1 = \rho_1$$

$$\phi_1 \rho_1 + \phi_2 = \rho_2$$

encontramos

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} \approx 0.5538$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \approx -0.1265$$

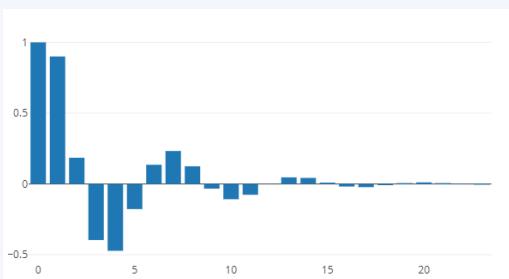
Las demás autocorrelaciones las encontramos con

$$\rho_j = 0.9\rho_{j-1} - 0.625\rho_{j-2}$$

Usando las fórmulas del tema 2 (pp.19-25), encontramos su función impulso respuesta

$$0.79^j [\cos(0.97j) - \frac{9}{13} \sin(0.97j)]$$

la cual converge a cero, oscilando periódicamente conforme $j \rightarrow \infty$.

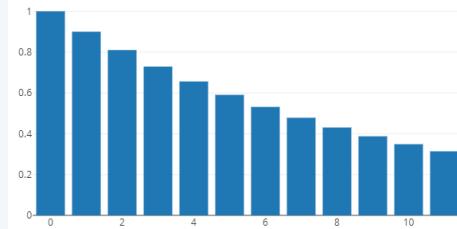


Ejemplo 3-7: AR(1) vs AR(2)

- Consideremos estos dos procesos autorregresivos:

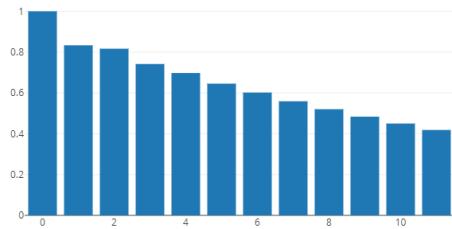
$$y_t = 0.9y_{t-1} + \epsilon_t$$

Autocorrelations



$$y_t = 0.5y_{t-1} + 0.4y_{t-2} + \epsilon_t$$

Autocorrelations



- Este caso ilustra que es sumamente difícil identificar el orden p de un proceso AR(p) a partir de su autocorrelograma.
- Para resolver este problema, a continuación estudiaremos la *autocorrelación parcial*.

Autocorrelación parcial

- La autocorrelación parcial mide la correlación *restante* entre y_t y y_{t-k} una vez que se han eliminado la influencia de $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$.

$$y_t = \underbrace{a_1^{(k)} y_{t-1} + a_2^{(k)} y_{t-2} + \cdots + a_{k-1}^{(k)} y_{t-k+1}}_{\text{"eliminamos" este efecto}} + a_k^{(k)} y_{t-k}$$

- Es decir, las primeras m autocorrelaciones parciales vienen de

$$\begin{aligned} y_t &= a_1^{(1)} y_{t-1} \\ y_t &= a_1^{(2)} y_{t-1} + a_2^{(2)} y_{t-2} \\ &\vdots \\ y_t &= a_1^{(m-1)} y_{t-1} + a_2^{(m-1)} y_{t-2} + \cdots + a_{m-1}^{(m-1)} y_{t-m+1} \\ y_t &= a_1^{(m)} y_{t-1} + a_2^{(m)} y_{t-2} + \cdots + a_{m-1}^{(m)} y_{t-m+1} + a_m^{(m)} y_{t-m} \end{aligned}$$

- Para encontrar el valor de $a_k^{(k)}$ basta con resolver

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1^{(k)} \\ a_2^{(k)} \\ a_3^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}$$

- En adelante, denotaremos la k -ésima correlación parcial por $\varphi(k) \equiv a_k^{(k)}$

- Comparando las ecuaciones del proceso AR(p) y de la autocorrelación parcial k :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

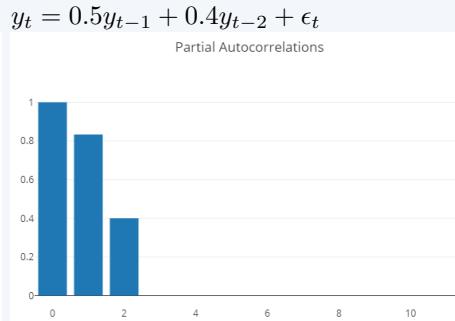
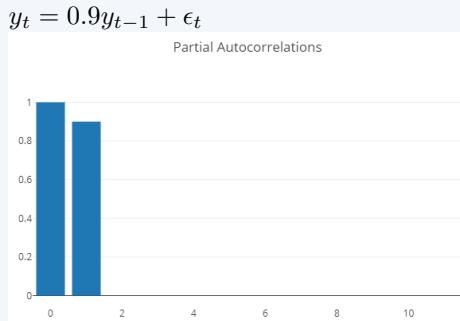
$$y_t = a_1^{(k)} y_{t-1} + a_2^{(k)} y_{t-2} + \cdots + a_{k-1}^{(k)} y_{t-k+1} + a_k^{(k)} y_{t-k}$$

vemos que

- si $k = p$, entonces $\varphi_k = \phi_p$
- si $k > p$, entonces $\varphi_k = 0$
- si $k = 1$, entonces $\varphi_1 = \rho_1$

Ejemplo 3-8: AR(1) vs AR(2), con autocorrelación parcial

- Consideremos de nuevo estos procesos autorregresivos:



- Ahora es muy sencillo identificar el orden p de los proceso AR.

Para encontrar la segunda autocorrelación parcial, resolvemos

$$\begin{bmatrix} a_1 \\ \varphi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \frac{1}{1 - \rho_1^2} \begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

de donde $\varphi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$.

- Para el proceso AR(1) sabemos que $\rho_k = \phi^k$, con lo que comprobamos que $\varphi_2 = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$.
- En un ejemplo anterior encontramos que para un proceso AR(2) se cumple

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} = \frac{5}{6} \quad \rho_2 = \phi_1 \rho_1 + \phi_2 = \frac{49}{60}$$

Al sustituir en la expresión para φ_2 comprobamos que

$$\varphi_2 = \frac{\frac{49}{60} - \frac{25}{36}}{1 - \frac{25}{36}} = 0.4 = \phi_2.$$

3.4 Proceso autorregresivo de media móvil ARMA(p,q)

Procesos ARMA

Definición: proceso ARMA

Sea $\{\epsilon_t\}$ un proceso ruido blanco; el proceso estocástico

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

con $\phi_p, \theta_q \neq 0$ es llamado un proceso ARMA(p,q).

- ARMA = AutoRegressive Moving Average, (autorregresivo media móvil)
- Los procesos ARMA son importantes porque todo proceso estacionario puede ser aproximado por un proceso ARMA.
- Similar a lo que encontramos con los procesos AR(p), si asumimos que es estacionario su media satisface la relación

$$\mu = c + \phi_1 \mu + \cdots + \phi_p \mu$$

- Por lo que el proceso puede escribirse sin el intercepto, si expresamos la variable y como desviación de su media

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Las fórmulas para la varianza y la autocovarianza se obtienen aplicando las definiciones del caso, pero tienden a ser más complicadas.

Estabilidad e invertibilidad de un proceso ARMA

El proceso $\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \cdots + \phi_p \tilde{y}_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$ puede expresarse en términos de polinomios de rezagos:

$$\begin{aligned} \tilde{y}_t - \phi_1 \tilde{y}_{t-1} - \cdots - \phi_p \tilde{y}_{t-p} &= \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ (1 - \phi_1 L - \cdots - \phi_p L^p) \tilde{y}_t &= (1 + \theta_1 L + \cdots + \theta_q L^q) \epsilon_t \\ \Phi(L) \tilde{y}_t &= \Theta(L) \epsilon_t \end{aligned}$$

Estabilidad

Si las raíces del polinomio $\Phi(z)$ están todas fuera del círculo unitario, el proceso es *estable*.

$$\tilde{y}_t = \Phi(L)^{-1} \Theta(L) \epsilon_t$$

Invertibilidad

Si las raíces del polinomio $\Theta(z)$ están todas fuera del círculo unitario, el proceso es *invertible*.

$$\epsilon_t = \Theta(L)^{-1} \Phi(L) \tilde{y}_t$$

Sobreparametrización de un proceso ARMA

- Supongamos que tenemos un proceso ARMA(p, q) $\Phi(L)y_t = \Theta(L)\epsilon_t$, en el cual los polinomios $\Theta(L)$ tiene una raíz en común. En este caso podemos

$$\begin{aligned} \Phi(L)y_t &= \Theta(L)\epsilon_t \\ (1 - rL)\Phi^*(L)y_t &= (1 - rL)\Theta^*(L)\epsilon_t \\ \Phi^*(L)y_t &= \Theta^*(L)\epsilon_t \end{aligned}$$

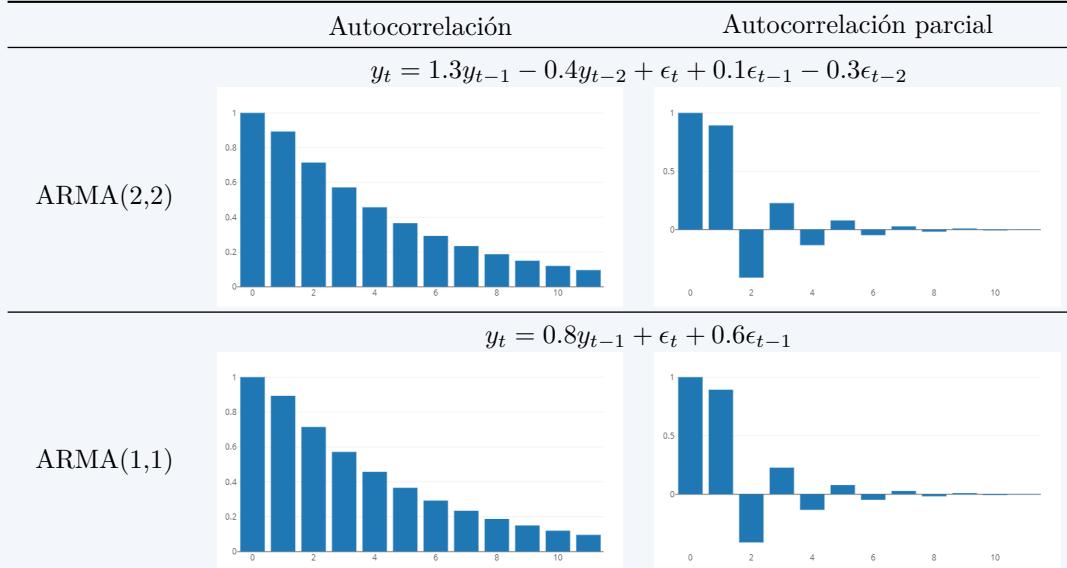
- Es decir, podemos representar el mismo proceso con un modelo ARMA(p-1, q-1).

- Decimos que:

- el modelo ARMA(p,q) está *sobreparametrizado*.
- el modelo ARMA(p-1, q-1) es una representación más *parsimoniosa* del proceso generador de datos.

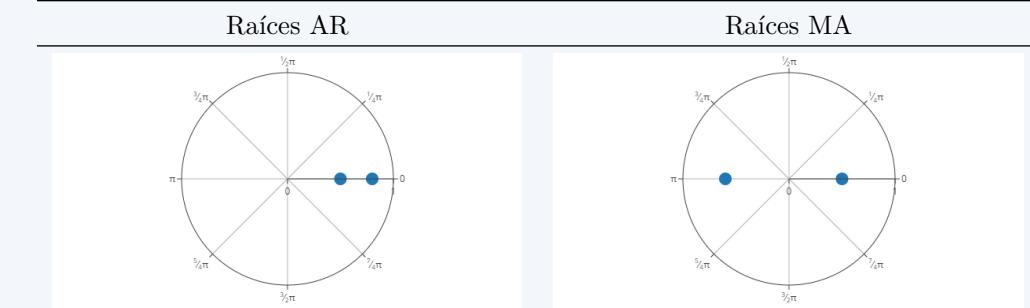
Ejemplo 3-9: Sobreparametrización

Consideremos estos dos procesos ARMA



¡Sus funciones ACF y PACF son idénticas!

- Las raíces de los polinomios de rezagos del proceso ARMA(2,2)



- Vemos que ambas tienen a $z = 0.5$ como una raíz.
- Al “eliminarla” de ambos polinomios, terminamos con el mismo proceso pero con menos parámetros.
- Esta representación ARMA(1, 1) es una versión más *parsimoniosa* del mismo proceso.

3.5 Estimación de modelos ARMA

Introducción (muy breve) al estimador de máxima verosimilitud

- Sea $f(y|\theta)$ la función de densidad conjunta de la variable $Y = [Y_1, \dots, Y_n]$. Entonces, para una muestra observada \mathbf{y} de esta distribución, la función del vector de parámetros θ definida por

$$L(\theta|\mathbf{y}) = f(\mathbf{y}|\theta)$$

se conoce como la *función de verosimilitud*.

- El *estimador de máxima verosimilitud* es el valor del vector de parámetros θ que maximiza la función de verosimilitud

$$\hat{\theta}_{\text{ML}} \equiv \underset{\theta}{\operatorname{argmax}} \underbrace{\mathcal{L}(\theta|\mathbf{y})}_{\ln L[(\theta|\mathbf{y})]}$$

- Es decir, $\hat{\theta}_{\text{ML}}$ es el parámetro que hace más plausible haber obtenido la muestra \mathbf{y} si la verdadera distribución era $f(y|\theta)$.

Ejemplo 3-10: Estimador máximo verosímil de una distribución exponencial

- Supongamos que tenemos una muestra $\{x_1, \dots, x_N\}$ de valores tomados de realizaciones independientes de una distribución exponencial con parámetro λ
- La función de densidad de una observación es $f(x|\lambda) = \lambda e^{-\lambda x}$
- La función de verosimilitud es la probabilidad conjunta de observar esta muestra:

$$L(\lambda|x_1, \dots, x_N) = \prod_{i=1}^N \lambda e^{-\lambda x_i}$$

o bien, tomando su logaritmo

$$\begin{aligned} \mathcal{L}(\lambda|x_1, \dots, x_N) &= \sum_{i=1}^N [\ln \lambda - \lambda x_i] \\ &= N \ln \lambda - \lambda \sum_{i=1}^N x_i \end{aligned}$$

- Para encontrar el máximo:

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=1}^N x_i = 0$$

- Por lo tanto, el estimador de máxima verosimilitud es:

$$\hat{\lambda}_{\text{ML}} = \frac{N}{\sum_{i=1}^N x_i} = \frac{1}{\bar{x}}$$

donde \bar{x} es el promedio simple de los datos.

Estimación de modelos ARMA

- Pensemos en un proceso ARMA estacionario como una distribución conjunta de Y_1, \dots, Y_T , donde

$$\mathbb{E}(Y_t) = \mu \quad \text{Cov}(Y_t, Y_{t-j}) = \gamma_j$$

- Entonces

$$\begin{aligned} \mathbb{E} \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} &= \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \mu \\ \text{Var} \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} &= \gamma_0 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \dots & 1 \end{bmatrix} = \Omega \end{aligned}$$

- Supongamos que tenemos una serie de tiempo débilmente estacionaria, con datos generados por el proceso ARMA descrito anteriormente.

$$\mathbf{y} = [y_1, y_2, \dots, y_T]'$$

- Si asumimos que el ruido blanco ϵ_t tiene una distribución normal, entonces el proceso ARMA tiene una distribución normal multivariada
- Su función de log-verosimilitud es

$$\mathcal{L}(\theta) = -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\Omega^{-1}| - \frac{1}{2} (\mathbf{y} - \mu)' \Omega^{-1} (\mathbf{y} - \mu)$$

donde

$$\theta = [c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q]$$

- En general, los modelos ARMA se estiman por el método de máxima verosimilitud.
- Para ello, se igualan a cero las derivadas de la función $\mathcal{L}(\theta)$ con respecto a cada uno de los parámetros presentes en θ .
- El resultado es un sistema de ecuaciones no lineales que carece de solución cerrada.
- Por ello, es necesario recurrir a métodos numéricos para resolver estos sistemas de ecuaciones.
- Por razones de tiempo, no cubrimos estos métodos en este curso. Para más detalles, consulte Greene (2012, capítulo 5) y Miranda y Fackler (2002, capítulos 3 y 4)

Estimación de modelos autorregresivos

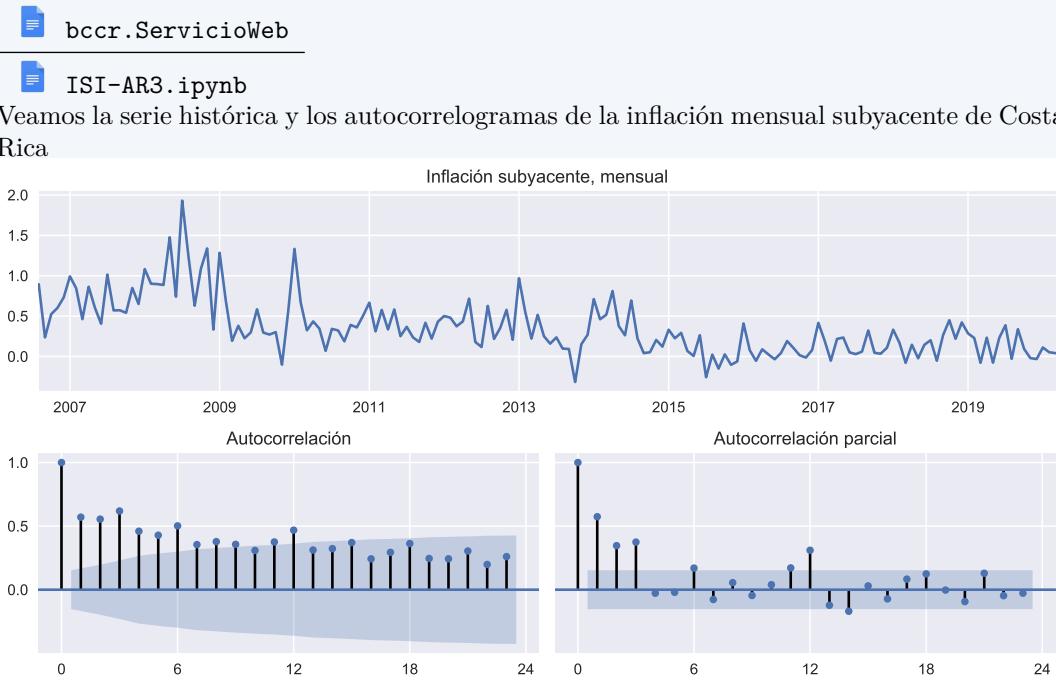
- Para el modelo AR(p)

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t \quad \text{Var}(\epsilon_t) = \sigma^2$$

es posible estimar los parámetros c, ϕ_1, \dots, ϕ_p por mínimos cuadrados ordinarios.

- El resultado es equivalente a un estimador de *máxima verosimilitud condicional*.
- Se llama así porque los primeros p valores de la serie se toman como dados.
- Si la muestra es grande, el resultado es equivalente al estimador de *máxima verosimilitud exacto*.
- El parámetro σ^2 se estima como el promedio de los cuadrados de los residuos de la regresión anterior.

Ejemplo 3-11: Estimando un modelo AR(3) para la inflación mensual



Consideremos este proceso AR(3) para modelar la inflación mensual subyacente

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t$$

la cual estimamos con datos mensuales (2006m08 a 2020m03, indicador 25725).

	coef	std err	z	P> z	[0.025	0.975]
const	0.3535	0.104	3.384	0.001	0.149	0.558
ar.L1.isi	0.2534	0.073	3.469	0.001	0.110	0.396
ar.L2.isi	0.1996	0.074	2.679	0.007	0.054	0.346
ar.L3.isi	0.3690	0.073	5.059	0.000	0.226	0.512

- El modelo ajustado es

$$\tilde{y}_t = 0.2534\tilde{y}_{t-1} + 0.1996\tilde{y}_{t-2} + 0.3690\tilde{y}_{t-3} + \epsilon_t$$

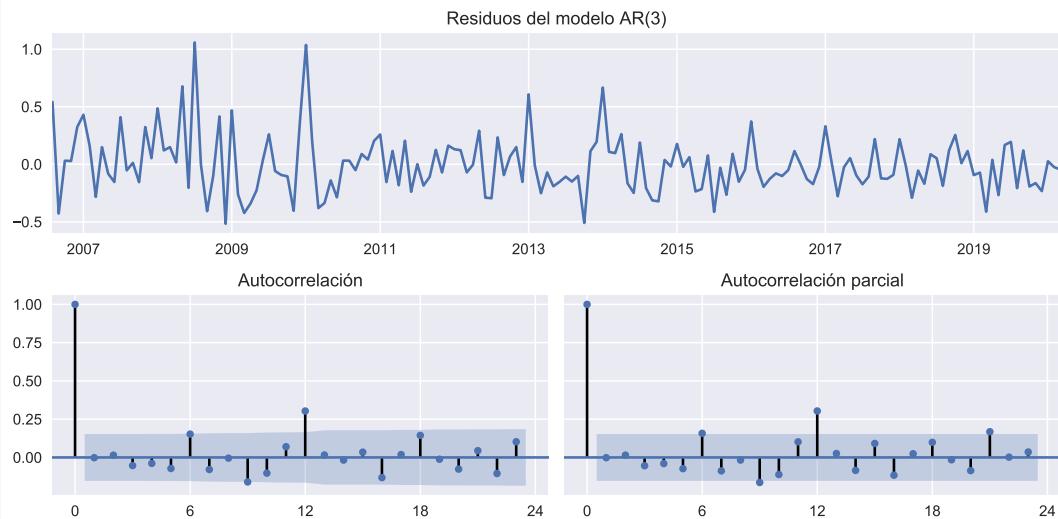
$$\tilde{y}_t = y_t - 0.3535$$

$$\epsilon_t \sim N(0, 0.2522^2)$$

- Las raíces inversas del polinomio $1 - 0.253 L - 0.200 L^2 - 0.369 L^3$ están dentro del círculo unitario, por lo que el proceso es estacionario.



Los residuos por la presencia de la regresión de un efecto parecen ruido estacional blanco, (rezago 12).



En Stata:

```

. arima isi, arima(3,0,0)
(setting optimization to BHHH)
Iteration 0: log likelihood = -7.3405352
Iteration 1: log likelihood = -7.3208124
Iteration 2: log likelihood = -7.3199769
Iteration 3: log likelihood = -7.3197271
Iteration 4: log likelihood = -7.3196206
(swapping optimization to BFGS)
Iteration 5: log likelihood = -7.3195774
Iteration 6: log likelihood = -7.3195452
Iteration 7: log likelihood = -7.3195446

ARIMA regression

Sample: 2006m8 - 2020m3
Number of obs = 164
Wald chi2(3) = 202.08
Prob > chi2 = 0.0000
Log likelihood = -7.319545



|             | isi      | Coef.    | Std. Err. | z     | P> z     | [95% Conf. Interval] |
|-------------|----------|----------|-----------|-------|----------|----------------------|
| <b>isi</b>  | _cons    | .3534968 | .1181878  | 2.99  | 0.003    | .121853 .5851406     |
| <b>ARMA</b> |          |          |           |       |          |                      |
| ar          |          |          |           |       |          |                      |
| L1.         | .2533589 | .0724494 | 3.50      | 0.000 | .1113607 | .3953571             |
| L2.         | .1995733 | .0514003 | 3.88      | 0.000 | .0988305 | .3003161             |
| L3.         | .3689968 | .0712193 | 5.18      | 0.000 | .2294095 | .5085841             |
| /sigma      | .2521808 | .0103002 | 24.48     | 0.000 | .2319929 | .2723688             |


```

Iteraciones del
método numérico

Selección de modelos

- En la estimación de máxima verosimilitud de un modelo ARMA(p,q) hay un supuesto implícito: que sabemos el orden del proceso, es decir, que sabemos el número “correcto” de rezagos p,q
- En la práctica, esto rara vez sucede.
- Tenemos una disyuntiva: Entre más rezagos incluyamos
 - “mejor” será el ajuste del modelo a los datos.
 - “peor” se vuelve la precisión de los parámetros que se estiman.
- La metodología de Box-Jenkins sugiere buscar modelos *parsimoniosos*.
- La parsimonia (usar tan pocos parámetros como sea necesario) tiene sus beneficios a la hora de hacer pronósticos.
- Muchos modelos estructurales complejos tienen un ajuste muy alto a la muestra en que se estiman, pero hacen pronósticos muy pobres fuera de la muestra.
- Sorprendentemente, modelos ARMA univariados sencillos pueden hacer mejores pronósticos.
- La idea es que entre más parámetros haya que estimar, más posibilidad hay de hacerlo mal.

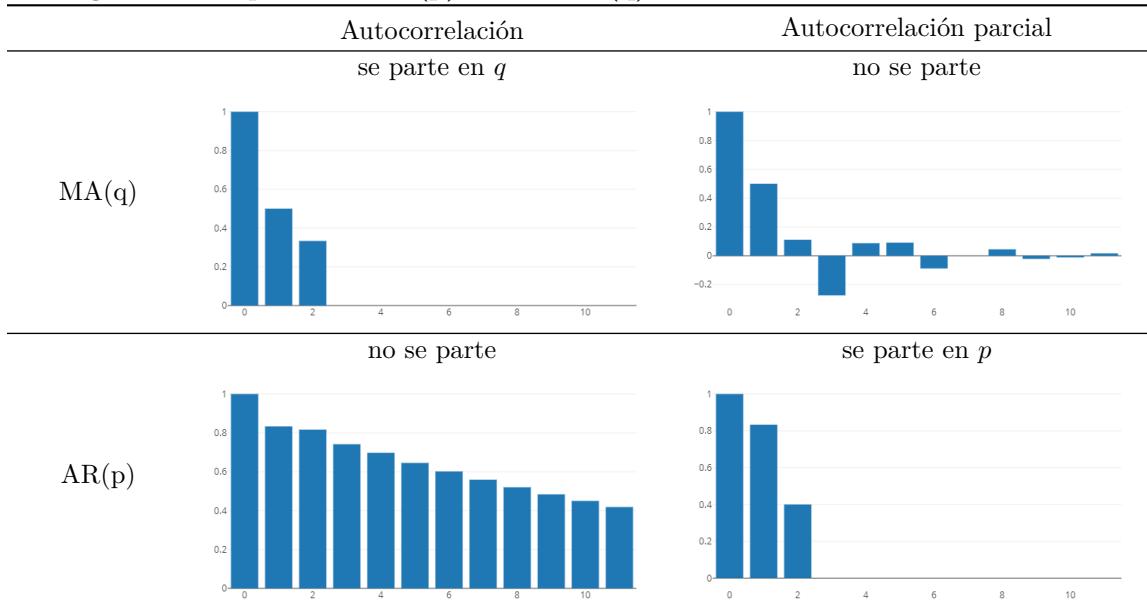
La filosofía de modelación de Box-Jenkins

El enfoque para pronosticar de Box-Jenkins consiste de cuatro etapas:

1. Transformar los datos (de ser necesario) para que el supuesto de estacionariedad sea razonable.

2. Adivinar valores pequeños de p y q para un modelo ARMA(p,q) que pueda describir la serie transformada.
3. Estimar los parámetros de $\phi(L)$ y $\theta(L)$.
4. Realizar análisis de diagnóstico para confirmar que el modelo es consistente con los datos observados.

Distinguiendo los procesos AR(p) de los MA(q)



Criterios de selección

- El autocorrelograma y el autocorrelograma parcial ayudan a reconocer series ARMA($p,0$) y ARMA($0,q$), pero no series ARMA(p,q) con $pq \neq 0$.
- Para esto, recurrimos a *criterios de selección*.
- Estos criterios tratan de resolver la disyuntiva de que a mayores valores de p, q , “mejor” será el ajuste pero “peor” la precisión de la estimación.
- Los criterios más usuales son el de Akaike (Akaike) y el de Bayes (BIC).
- Sean \mathcal{L} el máximo de la función log-verosimilitud, T el número de observaciones, y $K = p+q+2$ el número de parámetros estimado.
- Entonces

Criterio de información de Akaike

$$\text{AIC} = \frac{-2\mathcal{L}}{\text{“desajuste”}} + \frac{2K}{\text{penalización}}$$

Criterio de información de Bayes

$$\text{BIC} = \frac{-2\mathcal{L}}{\text{“desajuste”}} + \frac{\ln(T)K}{\text{penalización}}$$

- Se escoge la combinación p, q que minimiza el criterio de información.
- En la práctica, en ocasiones AIC y BIC escogen modelos distintos.

Ejemplo 3-12: Seleccionando p, q para un modelo ARMA de inflación

bccr.ServicioWeb

ISI-AR3.ipynb

En Stata, para el modelo AR(3) que estimamos en el ejemplo anterior:

Akaike's information criterion and Bayesian information criterion						
Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	T= 164	L= -7.319545	K= 5	24.63909	40.13842	

Calculamos los criterios de información:

$$\begin{aligned} \text{AIC} &= -2\mathcal{L} + 2K \\ &= -2 \times -7.3195 + 2 \times 5 = 24.63909 \end{aligned}$$

$$\begin{aligned} \text{BIC} &= -2\mathcal{L} + \ln(T)K \\ &= -2 \times -7.3195 + \ln(164) \times 5 = 40.13842 \end{aligned}$$

- Si calculamos los dos criterios para una combinación de valores p, q obtenemos

AIC

	q=0	q=1	q=2
p=0	128.02	89.62	83.77
p=1	64.31	30.74	31.92
p=2	46.27	32.21	34.44
p=3	24.64	26.51	28.32
p=4	26.50	28.43	25.10

BIC

	q=0	q=1	q=2
p=0	134.22	98.92	96.17
p=1	73.61	43.14	47.42
p=2	58.67	47.71	53.04
p=3	40.14	45.11	50.02
p=4	45.10	50.13	49.90

- Esto nos indica que la serie de tiempo debe modelarse como un proceso AR(3).

3.6 Pronósticos con modelos ARMA

Pronósticos con modelos ARMA

- Sea $y_{t+j|t}^*$ un pronóstico de y_{t+j} pasado en datos observados hasta t .
- Definimos el “mejor” pronóstico de este tipo como aquel que minimiza el *error cuadrático medio*

$$\text{MSE}\left(y_{t+j|t}^*\right) \equiv \mathbb{E}\left(y_{t+j} - y_{t+j|t}^*\right)^2$$

- Aunque no lo probamos aquí, resulta que el pronóstico con el menor error cuadrático medio es la esperanza de y_{t+j} condicional en los datos hasta t :

$$y_{t+j|t}^* = \mathbb{E}(y_{t+j}|y_t, y_{t-1}, \dots) \equiv \mathbb{E}_t(y_{t+j})$$

Pronosticando valores puntuales de la serie

- Es muy sencillo obtener de manera recursiva estos pronósticos para los modelos ARMA, siguiendo estos 3 pasos:
 1. Se escribe la ecuación ARMA de manera que y_t quede a la izquierda y todos los otros términos a la derecha.
 2. Se sustituye el índice t por $T + h$.
 3. En el lado derecho de la ecuación, se sustituyen:
 - observaciones futuras con sus pronósticos,
 - errores futuros con cero,
 - errores pasados con sus respectivos residuos.
- Empezando con $h = 1$, se repiten los pasos 2 y 3 para valores $h = 2, 3, \dots$ hasta que todos los pronósticos hayan sido calculados.

Para ilustrar el procedimiento, consideremos este proceso ARMA(1,2)

$$(1 - \phi L)\tilde{y}_t = (1 + \theta_1 L + \theta_2 L^2)\epsilon_t$$

- Paso 1: $\tilde{y}_t = \phi\tilde{y}_{t-1} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$
- Para $h = 1$:
$$\begin{aligned}\tilde{y}_{T+1} &= \phi\tilde{y}_T + \epsilon_{T+1} + \theta_1\epsilon_T + \theta_2\epsilon_{T-1} \\ \tilde{y}_{T+1|T}^* &= \phi\tilde{y}_T + \theta_1\hat{\epsilon}_T + \theta_2\hat{\epsilon}_{T-1}\end{aligned}$$
- Para $h = 2$:
$$\begin{aligned}\tilde{y}_{T+2} &= \phi\tilde{y}_{T+1} + \epsilon_{T+2} + \theta_1\epsilon_{T+1} + \theta_2\epsilon_T \\ \tilde{y}_{T+2|T}^* &= \phi\tilde{y}_{T+1|T}^* + \theta_2\hat{\epsilon}_T\end{aligned}$$
- Para $h = 3$:
$$\begin{aligned}\tilde{y}_{T+3} &= \phi\tilde{y}_{T+2} + \epsilon_{T+3} + \theta_1\epsilon_{T+2} + \theta_2\epsilon_{T+1} \\ \tilde{y}_{T+3|T}^* &= \phi\tilde{y}_{T+2|T}^*\end{aligned}$$
- Es fácil comprobar que, siguiendo este procedimiento, una vez que $h > p, h > q$ la ecuación de pronóstico será

$$\tilde{y}_{T+h}^* = \phi_1\tilde{y}_{T+h-1}^* + \cdots + \phi_p\tilde{y}_{T+h-p}^*$$

- Esto es una ecuación en diferencia de orden p , de la cual sabemos que sus raíces están dentro del círculo unitario.
- Por ello

$$\lim_{h \rightarrow \infty} \tilde{y}_{T+h}^* = 0 \quad \text{es decir} \quad \lim_{h \rightarrow \infty} y_{T+h}^* = \mu$$

- Esto nos dice que, para valores grandes de h , el pronóstico es simplemente la media del proceso.

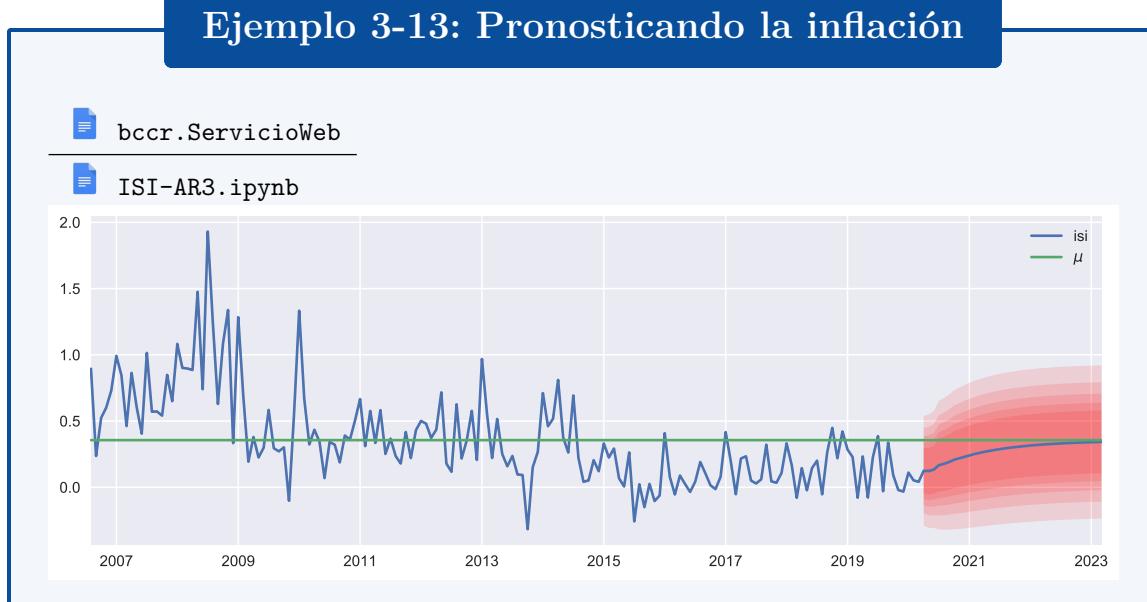
Cuantificando la incertidumbre de los pronósticos puntuales

- Para saber qué tan precisos esperamos que sean estos pronósticos, necesitamos cuantificar su error cuadrático medio

$$\begin{aligned}\text{MSE} \left(y_{t+j|t}^* \right) &\equiv \mathbb{E} \left(y_{t+j} - y_{t+j|t}^* \right)^2 \\ &= \mathbb{E} \left(y_{t+j} - \mathbb{E}_t (y_{t+j}) \right)^2 \\ &= \text{Var}_t (y_{t+j})\end{aligned}$$

- Es decir, necesitamos cuantificar la varianza de y_{t+j} condicional en los datos disponibles en t .
- Por razones de tiempo, no vamos a desarrollar estas fórmulas durante este curso.

Ejemplo 3-13: Pronosticando la inflación



¿Qué podría salir mal?

- En todas las fórmulas que hemos elaborado para los pronósticos hasta ahora, está implícito que conocemos los verdaderos valores de los parámetros.
- En la práctica, esos parámetros son estimados a partir de los datos.
- Tomemos por ejemplo un proceso AR(1):

$$\begin{array}{ll} y_{t+1} = \phi y_t + \epsilon_{t+1} & \text{(valor verdadero)} \\ y_{t+1|t}^* = \hat{\phi} y_t & \text{(pronóstico)} \\ \hline y_{t+1} - y_{t+1|t}^* & = (\phi - \hat{\phi}) y_t + \epsilon_{t+1} \\ \text{error de pronóstico} & \text{"sesgo"} \quad \text{innovación} \end{array}$$

Referencias del capítulo 3

- Becketti, Sean (2013). *Introduction to Time Series Using Stata*. 1^a ed. Stata Press. ISBN: 978-1-59718-132-7.
- Enders, Walter (2015). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Greene, William H. (2012). *Econometric Analysis*. 7^a ed. Prentice Hall. ISBN: 978-0-13-139538-1.
- Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
- Kirchgässner, Gebhard, Jürgen Wolters y Uwe Hassler (2013). *Introduction to Modern Time Series Analysis*. 2^a ed. Springer. ISBN: 978-3-642-33435-1.
- Miranda, Mario J. y Paul L. Fackler (2002). *Applied Computational Economics and Finance*. MIT Press. ISBN: 0-262-13420-9.

Capítulo 4

Modelos de volatilidad

En este capítulo

4.1	Introducción	69
4.2	El modelo ARCH	70
4.3	El modelo GARCH	82
4.4	Variantes del modelo GARCH	85

4.1 Introducción

La volatilidad de muchas series no es constante

- En modelos econométricos convencionales, se asume que la varianza del término de error es constante.
- Muchas series de tiempo económicas exhiben períodos de volatilidad inusualmente alta, seguidos por períodos de relativa tranquilidad.
- En tales circunstancias, el supuesto de *homoscedasticidad* es inapropiado.
- En ocasiones, uno puede estar interesado en pronosticar la *varianza condicional* de una serie.
- Consideremos el mercado accionario. Algunas veces el mercado es muy volátil, otras veces no.
- La volatilidad del retorno de las acciones determina el riesgo de las inversiones.
- En finanzas se tiene por cierto que el riesgo y el retorno están correlacionados positivamente
- Para hacer buenas inversiones, es crucial entender el riesgo apropiadamente.

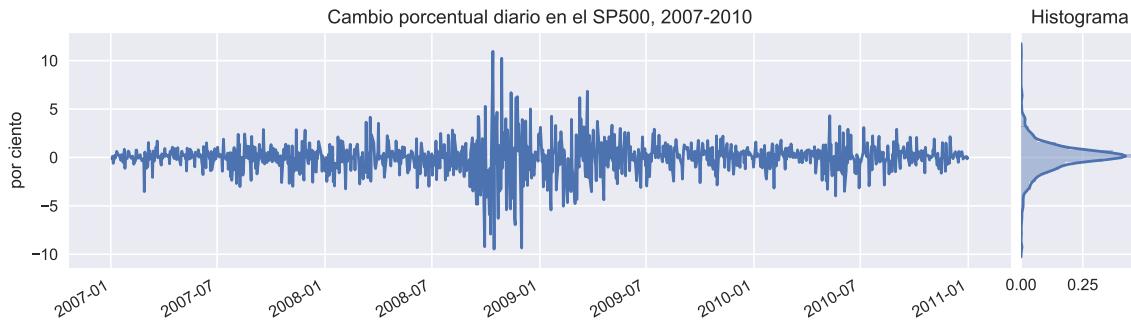


Figura 4.1: Volatilidad en el mercado accionario

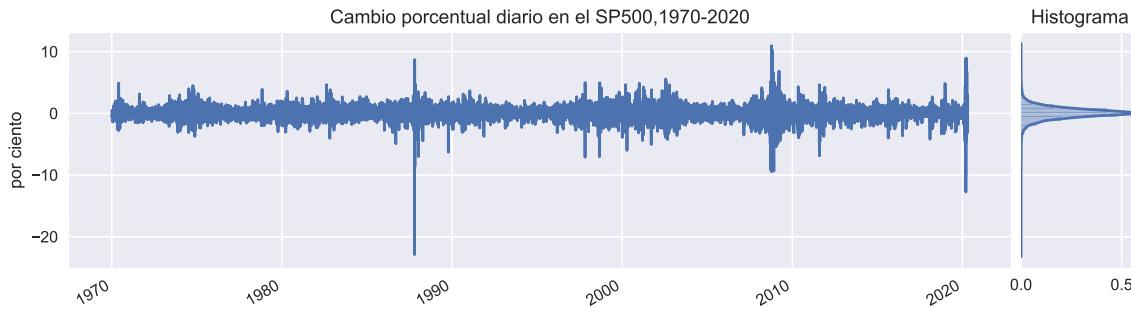
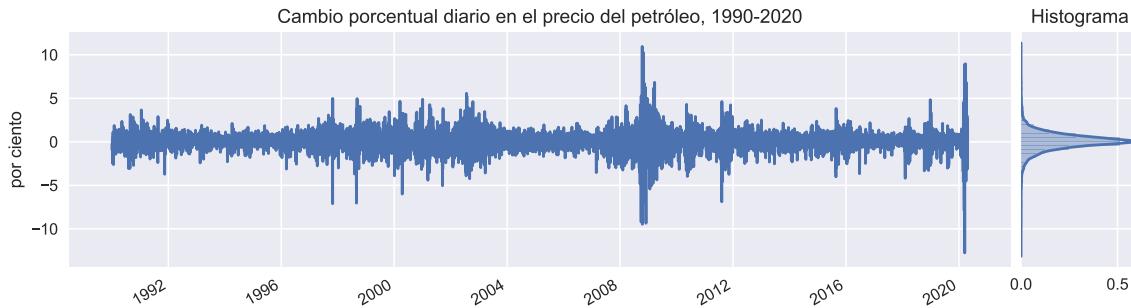


Figura 4.2: Volatilidad en el mercado petrolero



Pronosticando la varianza

- Un enfoque para pronosticar la varianza es introducir explícitamente una variable independiente que ayude a predecir la volatilidad.

- Por ejemplo $y_{t+1} = x_t \epsilon_{t+1}$

- Si $x_t = x_{t-1}$ es una constante observable en t , entonces $\{y_t\}$ es ruido blanco.

- De lo contrario, la varianza de y_{t+1} condicional en el valor observado x_t es

$$\text{Var}(y_{t+1} | x_t) = x_t^2 \sigma^2$$

- Si x_t tiene correlación serial positiva, entonces la varianza condicional de y_{t+1} también la tendrá.

4.2 El modelo ARCH

El modelo ARCH

- El modelo ARCH fue desarrollado por Engel (1982)
- Este trabajo le hizo co-ganador del Premio Nobel de Economía de 2003.
- En términos generales, todos los modelos ARCH (y GARCH, que estudiamos más adelante) consisten en dos ecuaciones
 1. una ecuación de la media, que describe la evolución de la variable de interés y_t ,
 2. una ecuación de la varianza, que describe la evolución de la varianza de y_t .
- En adelante, vamos a denotar por Ω_t todos los datos realizados hasta la fecha t .

El modelo ARCH(1)

ARCH(1)

El modelo ARCH(1) está definido por estas dos ecuaciones:

$$\begin{aligned} y_t &= c + \epsilon_t && \text{(media)} \\ \epsilon_t &= u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} && \text{(varianza)} \end{aligned}$$

donde $u_t \sim N(0, 1)$, $\alpha_0 > 0$, y $\alpha_1 > 0$.

- A continuación estudiamos los momentos condicionales y no condicionales del término de error ϵ_t .
- Luego vemos los momentos de la variable y_t .

Momentos del término de error

- La *media incondicional* del término de error es cero:

$$\begin{aligned} \mathbb{E} \epsilon_t &= \mathbb{E} \left[u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right] \\ &= \underbrace{\mathbb{E}[u_t]}_{=0} \mathbb{E} \left[\sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \right] = 0 \end{aligned}$$

- De manera similar, la *media condicional* en información previa también es cero:

$$\begin{aligned} \mathbb{E} [\epsilon_t | \Omega_{t-1}] &= \mathbb{E} \left[u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \mid \Omega_{t-1} \right] \\ &= \underbrace{\mathbb{E}[u_t | \Omega_{t-1}]}_{=0} \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} = 0 \end{aligned}$$

- Elevando al cuadrado la ecuación de la varianza es fácil calcular la *varianza no condicional* del error ϵ :

$$\begin{aligned}\epsilon_t^2 &= u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \\ \mathbb{E} \epsilon_t^2 &= \mathbb{E} [u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2)] \\ &= \mathbb{E}(u_t^2) \mathbb{E}(\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \\ &= \underbrace{\mathbb{E}(u_t^2)}_{=1} (\alpha_0 + \alpha_1 \underbrace{\mathbb{E} \epsilon_{t-1}^2}_{= \mathbb{E} \epsilon_t^2}) \\ \text{Var}(\epsilon_t) &= \frac{\alpha_0}{1 - \alpha_1}\end{aligned}$$

- De manera similar, la *varianza condicional* en la información disponible a $t-1$ es:

$$\begin{aligned}\mathbb{E} [\epsilon_t^2 | \Omega_{t-1}] &= \mathbb{E} [u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) | \Omega_{t-1}] \\ &= (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \underbrace{\mathbb{E} [u_t^2 | \Omega_{t-1}]}_{=1}\end{aligned}$$

- Por lo tanto:

$$\text{Var}[\epsilon_t | \Omega_{t-1}] = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$$

- Vemos que la varianza condicional tiene forma autorregresiva, de ahí el nombre del modelo ARCH (*AutoRegressive Conditional Heteroskedasticity*).
- Para futura referencia, resumimos en esta tabla los momentos que hemos calculado para la perturbación ϵ_t

	Incondicional	Condicional
Media	$\mathbb{E} \epsilon_t = 0$	$\mathbb{E} [\epsilon_t \Omega_{t-1}] = 0$
Varianza	$\text{Var} \epsilon_t = \frac{\alpha_0}{1 - \alpha_1}$	$\text{Var} [\epsilon_t \Omega_{t-1}] = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$

Momentos del proceso y

- La *media incondicional* de y_t es:

$$\begin{aligned}\mathbb{E}(y_t) &= \mathbb{E}(c + \epsilon_t) \\ &= c + \mathbb{E}(\epsilon_t) \\ &= c\end{aligned}$$

- mientras que su *varianza incondicional* es

$$\begin{aligned}\text{Var}(y_t) &= \text{Var}(c + \epsilon_t) \\ &= \text{Var}(\epsilon_t) \\ &= \frac{\alpha_0}{1 - \alpha_1}\end{aligned}$$

- La *media condicional* de y_t es:

$$\begin{aligned}\mathbb{E}(y_t | \Omega_{t-1}) &= \mathbb{E}(c + \epsilon_t | \Omega_{t-1}) \\ &= c + \mathbb{E}(\epsilon_t | \Omega_{t-1}) \\ &= c\end{aligned}$$

- mientras que su *varianza condicional* es

$$\begin{aligned}\text{Var}(y_t | \Omega_{t-1}) &= \text{Var}(c + \epsilon_t | \Omega_{t-1}) \\ &= \text{Var}(\epsilon_t | \Omega_{t-1}) \\ &= \text{Var}(\epsilon_t | \epsilon_{t-1}) \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2\end{aligned}$$

Ejemplo 4-1: Simulando un modelo ARCH(1)

 arch-simulaciones.do

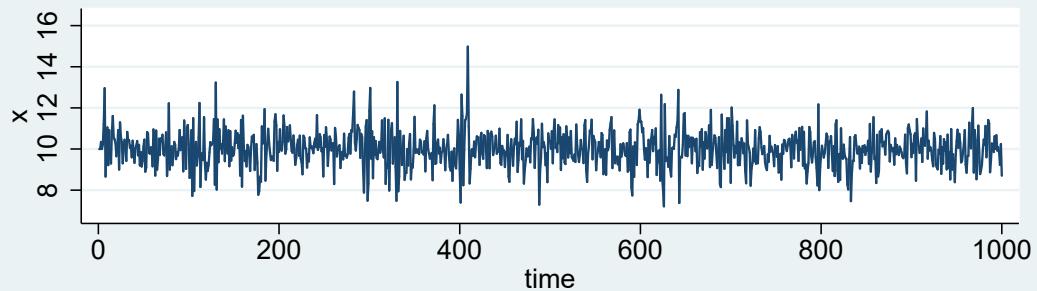
- En este y algunos de los próximos ejemplos vamos a simular procesos ARCH y GARCH.
- En todos ellos, se ejecuta este código al inicio

```
* Fijar parámetros de las simulaciones
set obs 1000
set seed 12345
gen time = _n
tsset time
gen u = rnormal(0,1)
```

Este código genera una realización del modelo

```
local c = 10
local alpha0 = 0.4
local alpha1 = 0.5
gen ex = 0

replace ex = u*(`alpha0' + `alpha1'*(L.ex^2))^(1/2) in 2/L
gen x = `c' + ex
tsline x
```



Para estimar el modelo

```
arch x, arch(1)

ARCH family regression
Sample: 1 - 1000          Number of obs = 1,000
Distribution: Gaussian     Wald chi2(.) =      .
Log likelihood = -1229.004 Prob > chi2 =      .
```

OPG

	x	Coef.	Std. Err.	z	P> z
x	_cons	9.993379	.0235107	425.06	0.000
ARCH	arch	.4658485	.0570226	8.17	0.000
	L1.				
	_cons	.4361722	.0295631	14.75	0.000

El modelo AR(1)-ARCH(1)

AR(1)-ARCH(1)

El modelo AR(1)-ARCH(1) está definido por estas dos ecuaciones:

$$y_t = c + \phi y_{t-1} + \epsilon_t \quad (\text{media})$$

$$\epsilon_t = u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} \quad (\text{varianza})$$

donde $u_t \sim N(0, 1)$, $|\phi| < 1$, $\alpha_0 > 0$, y $\alpha_1 > 0$.

- La ecuación de volatilidad es la misma de antes, por lo que los momentos de ϵ_t son los mismos del modelo ARCH(1).
- A diferencia del modelo ARCH(1), en esta especificación la media tiene una dinámica propia.
- Busquemos los momentos de la variable y_t .
- La *media incondicional* de y_t es:

$$\begin{aligned} \mathbb{E}(y_t) &= \mathbb{E}(c + \phi y_{t-1} + \epsilon_t) \\ &= c + \phi \underbrace{\mathbb{E}(y_{t-1})}_{=\mathbb{E} y_t} + \underbrace{\mathbb{E}(\epsilon_t)}_{=0} \\ &= c + \phi \mathbb{E}(y_t) \\ &= \frac{c}{1 - \phi} \end{aligned}$$

- mientras que la *media condicional* en Ω_{t-1} es

$$\begin{aligned} \mathbb{E}(y_t | \Omega_{t-1}) &= \mathbb{E}(c + \phi y_{t-1} + \epsilon_t | \Omega_{t-1}) \\ &= c + \phi \underbrace{\mathbb{E}(y_{t-1} | \Omega_{t-1})}_{=y_{t-1}} + \underbrace{\mathbb{E}(\epsilon_t | \Omega_{t-1})}_{=0} \\ &= c + \phi y_{t-1} \end{aligned}$$

- Por simplicidad, asumamos que $\mathbb{E}(y_t) = 0$

- Entonces su *varianza incondicional* es

$$\begin{aligned}\text{Var}(y_t) &= \mathbb{E}(y_t^2) \\ &= \mathbb{E}(\phi^2 y_{t-1}^2 + 2\phi y_{t-1} \epsilon_t + \epsilon_t^2) \\ &= \phi^2 \underbrace{\mathbb{E}(y_{t-1}^2)}_{=\text{Var}(y_t)} + 2\phi \mathbb{E}(y_{t-1} \epsilon_t) + \underbrace{\mathbb{E}(\epsilon_t^2)}_{=\alpha_0/(1-\alpha_1)} \\ (1-\phi^2) \text{Var}(y_t) &= 2\phi \underbrace{\mathbb{E}\left(y_{t-1} u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2}\right)}_{=0} + \frac{\alpha_0}{1-\alpha_1}\end{aligned}$$

por lo que

$$\text{Var}(y_t) = \frac{1}{1-\phi^2} \frac{\alpha_0}{1-\alpha_1}$$

- Por otra parte, su *varianza condicional* es

$$\begin{aligned}\text{Var}(y_t | \Omega_{t-1}) &= \text{Var}(c + \phi y_{t-1} + \epsilon_t | \Omega_{t-1}) \\ &= \text{Var}(\epsilon_t | \Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2\end{aligned}$$

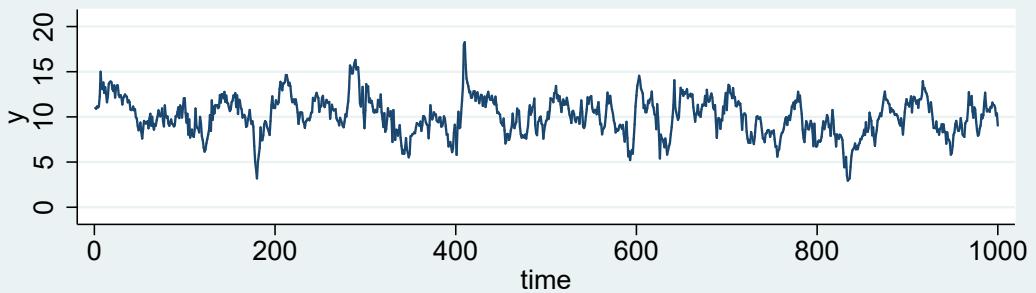
Ejemplo 4-2: Simulando un modelo AR(1)-ARCH(1)

 arch-simulaciones.do

Este código genera una realización del modelo

```
local c = 1      yt = 1 + 0.9yt-1 + εt          εt = ut √(0.4 + 0.5εt-12)
local phi1 = 0.9
local alpha0 = 0.4
local alpha1 = 0.5
gen ey = 0

replace ey = u*(`alpha0' + `alpha1'*(L.ey^2))^(1/2) in 2/L
gen y = 11
replace y = `c' + `phi1'*L.y + ey in 2/L
tsline y
```



Para estimar el modelo

<code>arch y L.y, arch(1)</code>							
ARCH family regression							
Sample: 2 - 1000		Number of obs = 999					
Distribution: Gaussian		Wald chi2(1) = 6915.39					
Log likelihood = -1227.867		Prob > chi2 = 0.0000					
<hr/>							
OPG							
		Coef.	Std. Err.	z P> z			
y							
y	L1.	.9127598	.0109761	83.16 0.000			
_cons		.8621232	.1121024	7.69 0.000			
<hr/>							
ARCH							
arch	L1.	.4683382	.0572297	8.18 0.000			
_cons		.4350871	.0296678	14.67 0.000			

El modelo ARCH(2)

ARCH(2)

El modelo ARCH(2) está definido por estas dos ecuaciones:

$$y_t = c + \epsilon_t \quad (\text{media})$$

$$\epsilon_t = u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2} \quad (\text{varianza})$$

donde $u_t \sim N(0, 1)$, $\alpha_i > 0$, $\alpha_1 + \alpha_2 < 1$.

- A continuación estudiamos los momentos condicionales y no condicionales del término de error ϵ_t .
- Luego vemos los momentos de la variable y_t .

Momentos del término de error

- La *media incondicional* del término de error es cero:

$$\begin{aligned} \mathbb{E} \epsilon_t &= \mathbb{E} \left[u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2} \right] \\ &= \underbrace{\mathbb{E}[u_t]}_{=0} \mathbb{E} \left[\sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2} \right] = 0 \end{aligned}$$

- De manera similar, la *media condicional* también es cero:

$$\begin{aligned} \mathbb{E} [\epsilon_t | \Omega_{t-1}] &= \mathbb{E} \left[u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2} \mid \Omega_{t-1} \right] \\ &= \underbrace{\mathbb{E}[u_t | \Omega_{t-1}]}_{=0} \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2} = 0 \end{aligned}$$

- Elevando al cuadrado la ecuación de la varianza es fácil calcular la *varianza no condicional* del error ϵ :

$$\begin{aligned}\epsilon_t^2 &= u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2) \\ \mathbb{E} \epsilon_t^2 &= \mathbb{E} [u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2)] \\ &= \mathbb{E} (u_t^2) \mathbb{E} (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2) \\ &= \underbrace{\mathbb{E} (u_t^2)}_{=1} (\alpha_0 + \alpha_1 \underbrace{\mathbb{E} \epsilon_{t-1}^2}_{= \mathbb{E} \epsilon_t^2} + \alpha_2 \underbrace{\mathbb{E} \epsilon_{t-2}^2}_{= \mathbb{E} \epsilon_t^2}) \\ \text{Var}(\epsilon_t) &= \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}\end{aligned}$$

- De manera similar, la *varianza condicional* en la información disponible a $t-1$ es:

$$\begin{aligned}\mathbb{E} [\epsilon_t^2 | \Omega_{t-1}] &= \mathbb{E} [u_t^2 (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2) | \Omega_{t-1}] \\ &= (\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2) \underbrace{\mathbb{E} [u_t^2 | \Omega_{t-1}]}_{=1}\end{aligned}$$

- Por lo tanto:

$$\text{Var}[\epsilon_t | \Omega_{t-1}] = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2$$

- Como $y_t = c + \epsilon_t$, vemos que las varianzas (condicional e incondicional) de y_t son iguales a las de ϵ_t .
- Entonces:

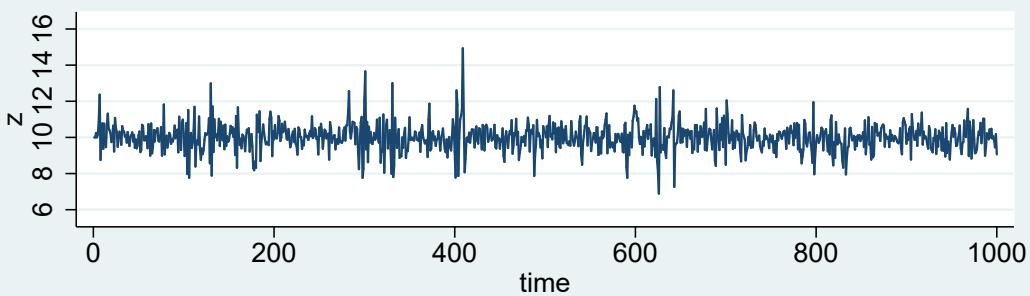
$$\begin{aligned}\text{Var} y_t &= \text{Var} \epsilon_t \\ &= \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}\end{aligned}$$

$$\begin{aligned}\text{Var}(y_t | \Omega_{t-1}) &= \text{Var}(\epsilon_t | \Omega_{t-1}) \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2\end{aligned}$$

Ejemplo 4-3: Simulando un modelo ARCH(2)

 arch-simulaciones.do
Este código genera una realización del modelo

```
z_t = 10 + epsilon_t
epsilon_t = u_t * sqrt(0.2 + 0.3 * epsilon_t^2 + 0.4 * epsilon_t^2)
local c = 10
local alpha0 = 0.2
local alpha1 = 0.3
local alpha2 = 0.4
gen ez = 0
replace ez = u*(`alpha0' + `alpha1'*(L.ez^2) + `alpha2'*(L.ez^2))^(1/2) in 3/L
gen z = 'c' + ez
tsline z
```



Para estimar el modelo

```
arch z, arch(1/2)
```

ARCH family regression

Sample: 1 - 1000 Number of obs = 1,000
 Distribution: Gaussian Wald chi2(.) = .
 Log likelihood = -992.1954 Prob > chi2 = .

		OPG			
		Coef.	Std. Err.	z	P> z
z	_cons	9.993861	.0170434	586.38	0.000
ARCH	arch				
	L1.	.658891	.0656346	10.04	0.000
	L2.	.0187866	.0271173	0.69	0.488
	_cons	.2137528	.0171361	12.47	0.000

El modelo ARCH(q)

ARCH(q)

El modelo ARCH(q) está definido por estas dos ecuaciones:

$$y_t = c + \epsilon_t \quad (\text{media})$$

$$\epsilon_t = u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2} \quad (\text{varianza})$$

- Para que la varianza sea estacionaria, se requiere que
 - $-1 < \alpha_i < 1$
 - $\sum_{i=1}^q \alpha_i < 1$
- En el caso del proceso ARCH(q), la *varianza condicional* es

$$\text{Var}(y_t | \Omega_{t-1}) = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2$$

- mientras que la *varianza incondicional* es

$$\text{Var}(y_t) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \dots - \alpha_q}$$

Determinando si los datos presentan ARCH

- ¿Cómo sabemos si los datos exhiben ARCH?
- Los cuadrados de los residuos son una estimación de la varianza, así que una forma de diagnosticar ARCH es estudiar la autocorrelación de los cuadrados de los residuos.
- Estudiaremos dos pruebas:
 1. La prueba de Ljung-Box
 2. La prueba de Engle basada en el multiplicador de Lagrange
- Ambas pruebas empiezan con los mismos dos pasos
 1. Se estima la ecuación de la media: se hace una regresión de y_t sobre sus rezagos o variables exógenas x_t
 2. Se investiga las propiedades de los residuos y de los cuadrados de los residuos

La prueba Q de Ljung-Box

- Recordemos que esta prueba sirve para determinar si una variable es ruido blanco.
- Se basa en la suma de los cuadrados de los primeros m coeficientes de autocorrelación, la cual debe ser “pequeña” si $\rho_1 = \rho_2 = \dots = \rho_m = 0$
- En este caso, calculamos el estadístico

$$Q = T(T + 2) \sum_{j=1}^m \frac{\hat{\rho}_j^2}{T - j} \xrightarrow{\text{asy}} \chi_{m-k}^2$$

a partir de las autocorrelaciones de los cuadrados de los residuos.

- En la fórmula, T es el número de observaciones (de la regresión, no de la serie original) y k el número de parámetros estimados en esa regresión (sin contar el intercepto).

La prueba de Engel

- Se estima una ecuación AR(m) para los residuos al cuadrado:

$$\hat{\epsilon}_t^2 = a_0 + a_1 \hat{\epsilon}_{t-1}^2 + a_2 \hat{\epsilon}_{t-2}^2 + \dots + a_m \hat{\epsilon}_{t-m}^2 + v_t$$

Test de Engel (1982)	
	¿Hay efectos ARCH en la serie de tiempo y_t ?
H_0	$a_1 = a_2 = \dots = a_m = 0$ (no hay ARCH)
Test	$\lambda = T' R^2 \xrightarrow{\text{asy}} \chi_m^2$
	Si $\lambda > \chi_m^2(1 - \alpha)$, rechazar H_0 con $100\alpha\%$ de significancia: la serie sí tiene efectos ARCH.

Encontrando el óptimo número de rezagos

- Tanto en la prueba de Ljung-Box como en la de Engle asumimos que conocíamos el orden m del proceso ARCH(m)
- En la práctica, eso no es así.
- Para determinar el valor de m , recurrimos de nuevo a los criterios de información:
 - criterio de información de Akaike
 - criterio de información bayesiano
- Escogemos el valor m que minimice estos criterios.
- ¿Y si no se pone de acuerdo? en tal caso el bayesiano escogerá un valor menor que el de Akaike:
 - usamos el bayesiano si preferimos una especificación más parsimoniosa,
 - usamos Akaike si nos preocupa incurrir en sesgo de variable omitida.

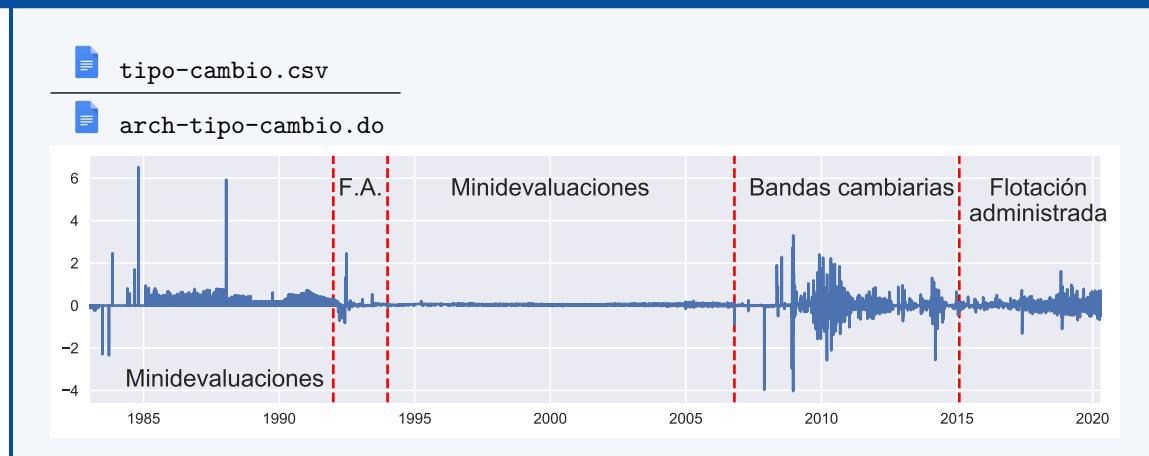
Estimando modelos ARCH

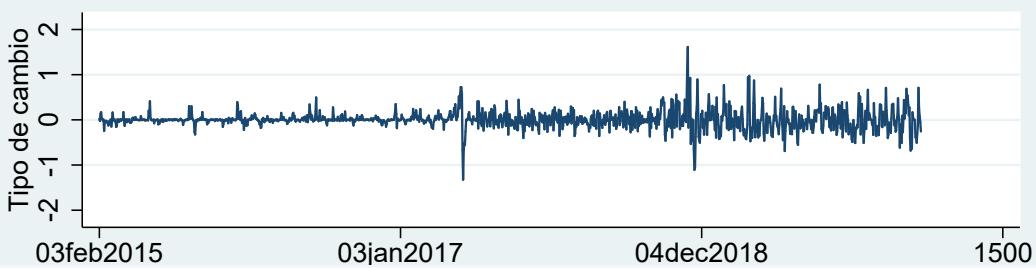
- En general, la ecuación de varianza de los modelos ARCH se pueden escribir como $\epsilon_t = u_t \sqrt{h_t}$, en los cuales siempre se asume que

$$\mathbb{E} u_t = 0 \quad \text{Var } u_t = 1 \quad \text{Cov}(u_t, u_{t-j}) = 0$$

- Hay varios métodos para estimar estos modelos:
 - el método de máxima verosimilitud,
 - el método de cuasi-máxima verosimilitud,
 - el método generalizado de momentos.
- Además, las estimaciones de máxima verosimilitud pueden asumir distintas distribuciones del ruido blanco u_t .

Ejemplo 4-4: Un modelo ARCH del tipo de cambio del colón/dólar



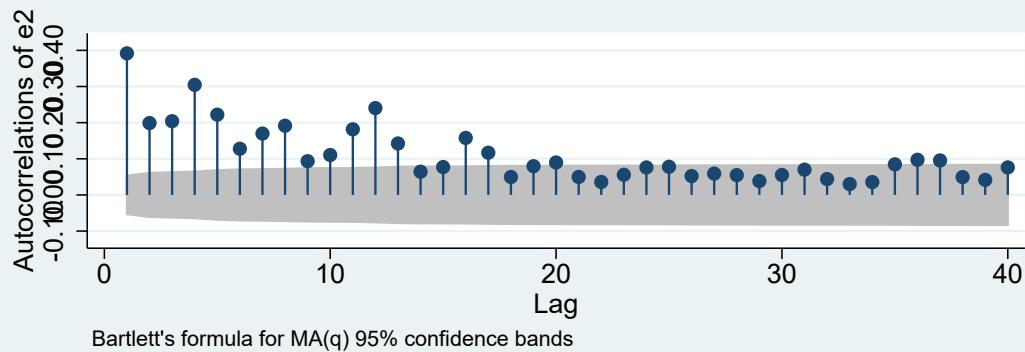


Calculamos los cuadrados de los residuos de la ecuación de la media:

```

1 quietly regress tc
2 predict e, resid
3 gen e2 = e^2
4
5 ac e2 /* autocorrelograma*/

```



Sin importar cuántos rezagos usamos, la prueba del multiplicador de lagrange nos dice que sí hay efectos ARCH.

```

estat archlm, lags(1/12)

```

IM test for autoregressive conditional heteroskedasticity (ARCH)

lags(p)	chi2	df	Prob > chi2
1	209.419	1	0.0000
2	212.541	2	0.0000
3	231.625	3	0.0000
4	283.783	4	0.0000
5	284.577	5	0.0000
6	284.471	6	0.0000
7	291.968	7	0.0000
8	293.769	8	0.0000
9	298.060	9	0.0000
10	301.888	10	0.0000
11	310.374	11	0.0000
12	322.986	12	0.0000

H0: no ARCH effects vs. H1: ARCH(p) disturbance

Los criterios de Akaike y el bayesiano coiciden en especificar un modelo ARCH(10).

```

forvalues lags=1/11{
    quietly arch tc, arch(1`lags')
}

```

```

quietly estat ic
matrix temp = r(S)
display "lags=" `lags' " AIC=" temp[1,5] " BIC=" temp[1,6]
}

lags = 1      AIC = -990.60684    BIC = -974.95011
lags = 2      AIC = -1218.8708    BIC = -1197.9952
lags = 3      AIC = -1276.3479    BIC = -1250.2534
lags = 4      AIC = -1315.5608    BIC = -1284.2473
lags = 5      AIC = -1361.0546    BIC = -1324.5222
lags = 6      AIC = -1378.8952    BIC = -1337.144
lags = 7      AIC = -1378.8461    BIC = -1331.8759
lags = 8      AIC = -1381.6959    BIC = -1329.5068
lags = 9      AIC = -1401.3779    BIC = -1343.9699
lags = 10     AIC = -1420.0086    BIC = -1357.3817
lags = 11     AIC = -1418.071     BIC = -1350.2252

```

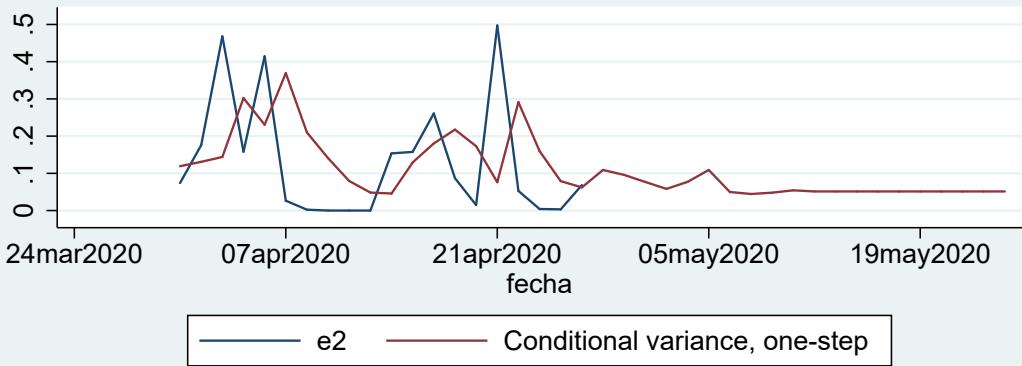
Estimamos el modelo ARCH con el número óptimo de rezagos y pronosticamos la varianza del siguiente mes:

```

arch tc, arch(1/10)
tsappend, add(20) /* ampliamos la muestra*/

predict varhat, variance /* pronóstico */
tsline e2 varhat in -40/1

```



4.3 El modelo GARCH

De ARCH a GARCH

- Los modelos ARCH pueden capturar muchas de las características de datos financieros, pero para lograrlo pueden necesitar muchos rezagos en la ecuación de varianza.
- En 1986 Bollerslev encontró una solución a este problema vía una generalización del modelo ARCH.
- Un modelo GARCH(p,q) puede reflejar un ARCH(∞) usando pocos parámetros.
 - Antes de generalizar el modelo ARCH(q), reescribamos la ecuación de la varianza así

$$\epsilon_t = u_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2} = \sigma_t u_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2$$

- Vemos que la varianza condicional σ_t^2 de un modelo ARCH(q) es similar a un proceso MA(q).
- El modelo GARCH(p,q) se obtiene agregando p rezagos de la varianza condicional al proceso ARCH(q):

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

El modelo GARCH(1,1)

GARCH(1,1)

El modelo GARCH(1,1) está definido por estas tres ecuaciones:

$$\begin{aligned} y_t &= c + \epsilon_t \\ \epsilon_t &= u_t \sigma_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned}$$

- Notemos lo parecida que es esta formulación de la varianza condicional a un proceso ARMA(1,1).
- Podemos escribirla así la varianza condicional

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\ \sigma_t^2 - \beta \sigma_{t-1}^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \\ (1 - \beta L) \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \\ \sigma_t^2 &= (1 - \beta L)^{-1} (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) \\ &= \frac{\alpha_0}{1 - \beta} + (1 + \beta L + \beta^2 L^2 + \dots) \alpha_1 \epsilon_{t-1}^2 \\ &= \frac{\alpha_0}{1 - \beta} + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \beta \epsilon_{t-2}^2 + \alpha_1 \beta^2 \epsilon_{t-3}^2 + \dots \end{aligned}$$

siempre y cuando $|\beta| < 1$.

- Vemos que el proceso GARCH(1,1) es equivalente a un proceso ARCH(∞).
- Esto permite al GARCH capturar procesos muy complejos sin necesidad de estimar muchísimos parámetros.

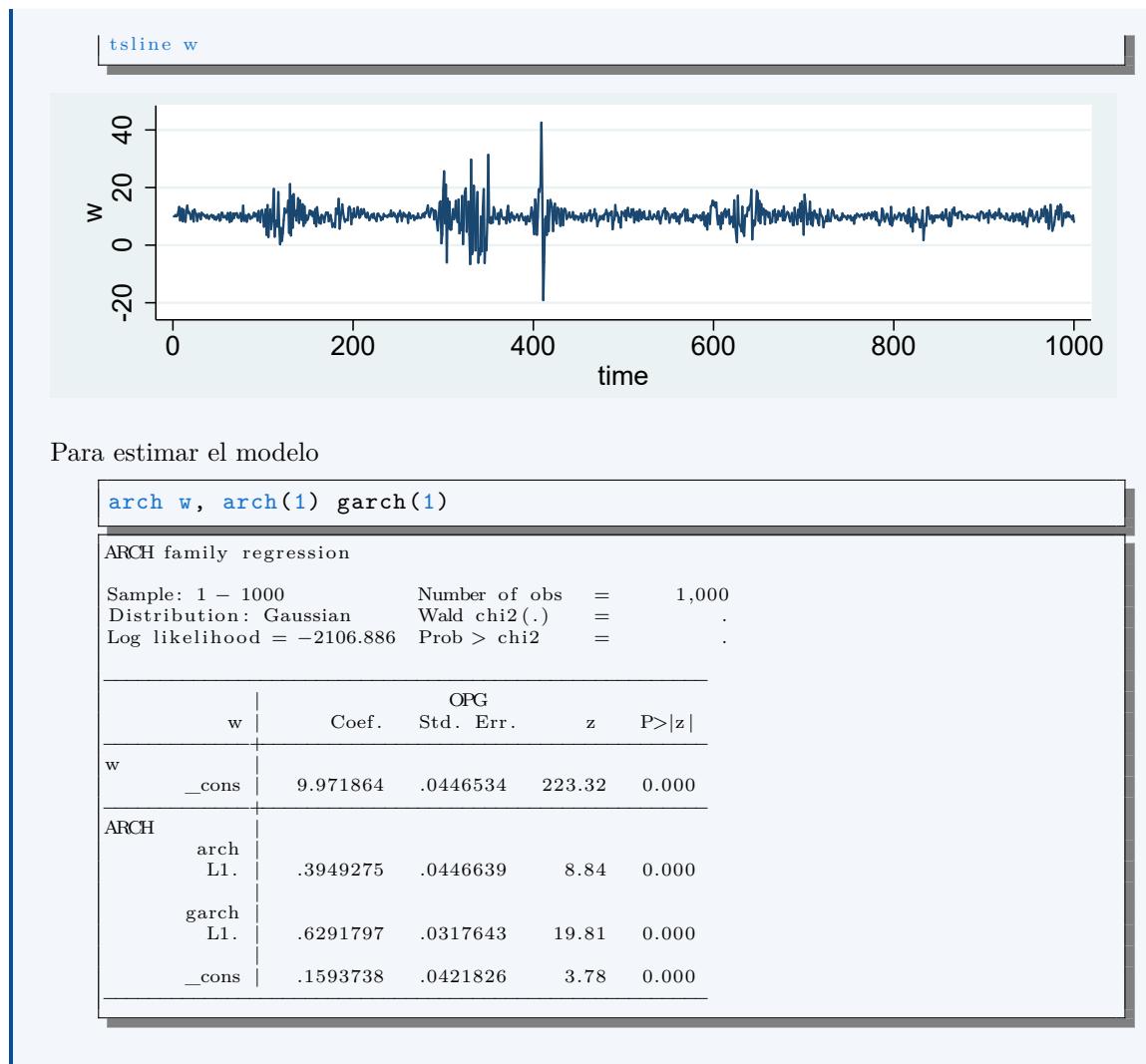
Ejemplo 4-5: Simulando un modelo GARCH(1,1)



Este código genera una realización del modelo

```
w_t = 10 + ε_t           ε_t = u_t σ_t           σ_t² = 0.2 + 0.4ε_{t-1}² + 0.6σ_{t-1}²
local c = 10
local alpha0 = 0.2
local alpha1 = 0.4
local beta1 = 0.6
gen ew = 0
gen sigma2 = 1

forvalues i=2/`=_N`{
    replace sigma2 = `alpha0' + `alpha1'*(`L.ew'^2) + `beta1'*(`L.sigma2') in `i'
    replace ew = `u*sqrt(sigma2)' in `i'
}
gen w = `c' + ew
```



El modelo GARCH(p,q)

GARCH(p,q)

El modelo GARCH(p,q) está definido por estas tres ecuaciones:

$$y_t = c + \epsilon_t$$

$$\epsilon_t = u_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2$$

- En la práctica, es inusual requerir más de dos rezagos ARCH y GARCH (i.e., $p \leq 2$ y $q \leq 2$).

Ejemplo 4-6: Un modelo GARCH del tipo de cambio del colón/dólar

tipo-cambio.csv

arch-tipo-cambio.do

Los criterios de Akaike y el bayesiano coinciden en especificar un modelo GARCH(2,3).

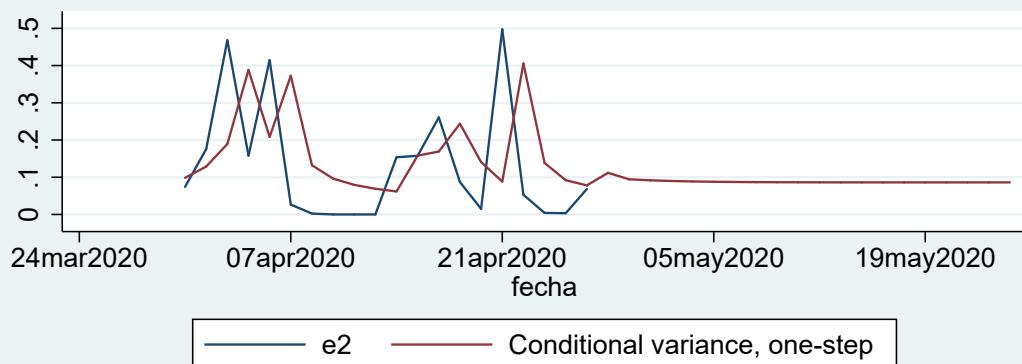
```
forvalues p=0/2{
forvalues q=1/4{
    if `p'==0{
        quietly arch tc, arch(1/`q')
    }
    else {
        quietly arch tc, arch(1/`q') garch(1/`p')
    }
    quietly estat ic
    matrix temp = r(S)
    display "p = " `p' " q = " `q' " AIC = " temp[1,5] " BIC = " temp[1,
    6]
}
}
```

```
p = 0 q = 1 AIC = -990.60684 BIC = -974.95011
p = 0 q = 2 AIC = -1218.8708 BIC = -1197.9952
p = 0 q = 3 AIC = -1276.3479 BIC = -1250.2534
p = 0 q = 4 AIC = -1315.5608 BIC = -1284.2473
p = 1 q = 1 AIC = -1415.7106 BIC = -1394.835
p = 1 q = 2 AIC = -1448.2148 BIC = -1422.1202
p = 1 q = 3 AIC = -1488.5201 BIC = -1457.2067
p = 1 q = 4 AIC = -1489.3458 BIC = -1452.8134
p = 2 q = 1 AIC = -1424.5417 BIC = -1398.4472
p = 2 q = 2 AIC = -1493.9636 BIC = -1462.6502
p = 2 q = 3 AIC = -1500.0086 BIC = -1463.4762
p = 2 q = 4 AIC = -1459.4402 BIC = -1417.6889
```

Estimamos el modelo GARCH con el número óptimo de rezagos y pronosticamos la varianza del siguiente mes:

```
arch tc, arch(1/3) garch(1/2)

predict varhat2, variance
tsline e2 varhat2 in -40/1
```



Aunque no mostramos más detalles acá, podemos también estimar modelos AR-GARCH. Por ejemplo este modelo AR(1)-GARCH(2,2) del tipo de cambio

```
arch tc L.tc, arch(1/2) garch(1/2)
```

Sample: 04feb2015 – 25apr2020 Number of obs = 1,364				
Distribution: Gaussian Wald chi2(1) = 96.45				
Log likelihood = 791.6779 Prob > chi2 = 0.0000				
tc	Coef.	Std. Err.	z	P> z
L1.tc	.31708	.0322866	9.82	0.000
_cons	.0020962	.0028756	0.73	0.466
ARCH				
L1.arch	.4853867	.0299791	16.19	0.000
L2.arch	-.4833929	.0299363	-16.15	0.000
L1.garch	1.453636	.0334281	43.49	0.000
L2.garch	-.454845	.033364	-13.63	0.000
_cons	-1.84e-06	1.41e-06	-1.31	0.192

4.4 Variantes del modelo GARCH

Variantes del modelo GARCH

- A la fecha se ha realizado mucha investigación para mejorar la capacidad de ajuste de los modelos GARCH.
- Las mejoras se logran principalmente cambiando las restricciones sobre los parámetros del modelo.
- Existen muchísimas variantes del modelo GARCH, entre ellas:
 - GARCH-t, que asume que el ruido blanco sigue una distribución *t*-Student en vez de una normal estándar.
 - GARCH-M, que incluye la varianza como una variable explicativa de la media.
 - GJR-GARCH, E-GARCH, T-GARCH, para modelar respuestas asimétricas a los shocks.
 - I-GARCH, para procesos que tienen una varianza con raíz unitaria.

El modelo GARCH-M

- En finanzas se tiene por hecho que mayores riesgos solo se asumen si se esperan mayores retornos.
- Esto sugiere una modificación del modelo GARCH: incluir la varianza del proceso en la ecuación del retorno esperado.
- Así, por ejemplo, puede especificarse el modelo

$$\begin{aligned}y_t &= c + \gamma x_t + \lambda \sigma_t + \epsilon_t \\ \epsilon_t &= \sigma_t u_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2\end{aligned}$$

- Este modelo puede generalizarse aún más, incluyendo además rezagos de la varianza en la ecuación de y_t .

Referencias del capítulo 4

- Enders, Walter (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
- Levendis, John D. (2018). *Time Series Econometrics. Learning Through Replication*. Springer. ISBN: 978-3-319-98281-6.

Capítulo 5

Modelos de tendencia

En este capítulo

5.1	Introducción	87
5.2	Modelos de series con tendencia	88
5.3	Removiendo una tendencia	93
5.4	Series con raíz unitaria y procesos ARIMA	95
5.5	Detectando raíces unitarias	99
5.6	Otro método para remover una tendencia	108

5.1 Introducción

Series con tendencia

- Hasta ahora, para cada uno de los procesos que hemos estudiado su valor esperado es constante.
- Sin embargo, muchas de las series que estudiamos en la práctica tienen una tendencia.
- En tales casos, un modelo de proceso estacionario (como ARMA) no es apropiado para describir la serie.
- Por ejemplo:



Componentes (no observados) de una serie

- Para esta lección, es útil imaginar que una serie de tiempo consiste de dos componentes distintos:

$$y_t = \text{tendencia}_t + \text{componente_estacionario}_t$$

- En el tema 3 del curso aprendimos que una serie estacionaria puede modelarse como un proceso ARMA.
- Ahora en este tema aprenderemos:
 - cómo se modela una tendencia,
 - cómo se determina si una serie tiene tendencia,
 - cómo extraer una tendencia.

5.2 Modelos de series con tendencia

Series *integradas*

- A las series que tienen tendencia se les conoce como *series integradas*
- ¿Pero por qué integradas?
- Supongamos por un momento que $y(t)$ es una variable cuyo valor depende del momento t (continuo) en que se observa.
- Supongamos también que y cambia a una tasa constante a y que su valor inicial es $y(0)$.
- ¿A qué es igual $y(t)$?

De los supuestos anteriores:

$$\begin{aligned}\frac{dy}{d\tau} &= a \\ dy &= a d\tau\end{aligned}$$

integrando ambos lados

$$\begin{aligned}\int_0^t dy &= \int_0^t a d\tau \\ y(\tau)|_0^t &= a\tau|_0^t \\ y(t) - y(0) &= at\end{aligned}$$

por lo tanto

$$y(t) = y(0) + at$$

- Supongamos ahora que el tiempo se mide en unidades discretas, que el valor inicial es y_0 y que la variable crece una cantidad fija a cada período.

- Entonces

$$\Delta y_\tau = a \underbrace{\Delta \tau}_{=1} = a$$

integrando ambos lados

$$\begin{aligned} \sum_{\tau=1}^t \Delta y_\tau &= \sum_{\tau=1}^t a \\ \frac{\Delta y_1}{y_1 - y_0} + \frac{\Delta y_2}{y_2 - y_1} + \frac{\Delta y_3}{y_3 - y_2} + \cdots + \frac{\Delta y_t}{y_t - y_{t-1}} &= at \\ y_t - y_0 &= at \end{aligned}$$

por lo tanto

$$y_t = y_0 + at$$

- Este resultado lo podemos obtener de manera equivalente con el operador de rezagos:

$$\begin{aligned} \Delta y_\tau &= a \\ (1 + L + L^2 + \cdots + L^{t-1}) \Delta y_t &= (1 + L + L^2 + \cdots + L^{t-1}) a \\ (1 + L + L^2 + \cdots + L^{t-1}) (1 - L) y_t &= \underbrace{a + a + \cdots + a}_{t \text{ veces}} \\ (1 - L^t) y_t &= at \\ y_t - y_0 &= at \end{aligned}$$

- Vemos que si $\Delta y_t = a$, es decir, la serie **cambia una cantidad fija** a cada período, entonces $y_t = y_0 + at$ tiene una *tendencia determinística*.
- Supongamos en contraposición que el **valor esperado del cambio** es constante, en particular

$$\Delta y_\tau = a + \epsilon_\tau$$

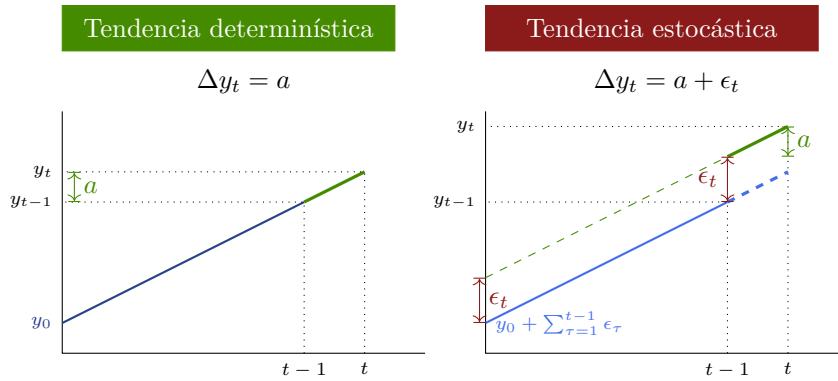
donde ϵ_τ es ruido blanco.

- Entonces, integrando

$$\begin{aligned} \sum_{\tau=1}^t \Delta y_\tau &= \sum_{\tau=1}^t (a + \epsilon_\tau) \\ y_t - y_0 &= at + \sum_{\tau=1}^t \epsilon_\tau \qquad \Rightarrow y_t = \left(y_0 + \sum_{\tau=1}^t \epsilon_\tau \right) + at \end{aligned}$$

- Decimos que la serie tiene una *tendencia estocástica*.

Tendencias determinísticas versus estocásticas



Tendencias cuadráticas

- Es sencillo generalizar el procedimiento para obtener tendencias cuadráticas.
- Si el cambio en la serie es lineal en vez de constante:

$$\begin{aligned}\Delta y_{\tau} &= a + b\tau \\ \sum_{\tau=1}^t \Delta y_{\tau} &= \sum_{\tau=1}^t (a + b\tau) = at + b \sum_{\tau=1}^t \tau \\ y_t - y_0 &= at + b \frac{t(t+1)}{2} \\ y_t &= y_0 + \frac{2a+b}{2}t + \frac{b}{2}t^2\end{aligned}$$

Serie estacionaria alrededor de una tendencia (TS)

- Supongamos que $\omega_t = A(L)\epsilon_t$ es una serie estacionaria, tal que

$$\mathbb{E} \omega_t = 0 \quad \text{Var } \omega_t = \sigma_{\omega}^2$$

- Definimos una serie

estacionaria alrededor de una tendencia

$$y_t = \underbrace{y_0 + at}_{\text{tendencia}} + \underbrace{\omega_t}_{\text{estacionario}}$$

- Vemos que

$$\mathbb{E} y_t = y_0 + at \quad \text{Var } y_t = \sigma_{\omega}^2$$

- Por otra parte, su autocovarianza es:

$$\begin{aligned}\text{Cov}(y_t, y_{t-s}) &= \mathbb{E}[(y_t - y_0 - at)(y_{t-s} - y_0 - a(t-s))] \\ &= \mathbb{E}[\omega_t \omega_{t-s}] \\ &= \text{Cov}[\omega_t, \omega_{t-s}]\end{aligned}$$

- Es decir, la media de y_t no es estacionaria (porque depende de t), pero su varianza-covarianza sí es estacionaria (idéntica a la del proceso ω_t).
- Por ello, el correlograma de un proceso estacionario alrededor de una tendencia es similar al correlograma de un proceso estacionario.

Caminata aleatoria

- El proceso de caminata aleatoria es muy utilizado en la teoría económica y financiera, en especial en la formulación de la hipótesis de mercado eficiente.
- Definimos la serie

caminata aleatoria

$$\Delta y_t = \epsilon_t \quad \text{o bien} \quad y_t = y_{t-1} + \epsilon_t$$

- Integrando el proceso sabemos que

$$y_t = y_0 + \sum_{\tau=1}^t \epsilon_\tau$$

- Consideremos y_0 como dado, y calculemos los momentos condicionales de la caminata aleatoria:

$$\begin{aligned} \mathbb{E} y_t &= \mathbb{E} \left[y_0 + \sum_{\tau=1}^t \epsilon_\tau \right] = y_0 \\ \text{Var } y_t &= \text{Var} \left[y_0 + \sum_{\tau=1}^t \epsilon_\tau \right] \\ &= \text{Var} [\epsilon_1 + \epsilon_2 + \cdots + \epsilon_t] = t\sigma^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(y_t, y_{t-s}) &= \mathbb{E} [(y_t - y_0)(y_{t-s} - y_0)] \\ &= \mathbb{E} [(\epsilon_1 + \cdots + \epsilon_{t-s} + \cdots + \epsilon_t)(\epsilon_1 + \cdots + \epsilon_{t-s})] \\ &= (t-s)\sigma^2 \end{aligned}$$

$$\begin{aligned} \rho_{t,s} &= \frac{\text{Cov}(y_t, y_{t-s})}{\sqrt{\text{Var } y_t \text{ Var } y_{t-s}}} = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2(t-s)\sigma^2}} \\ &= \sqrt{\frac{t-s}{t}} = \sqrt{1 - \frac{s}{t}} \end{aligned}$$

Caminata aleatoria con deriva

- El proceso de caminata aleatoria con deriva es similar a la caminata aleatoria:

caminata aleatoria con deriva

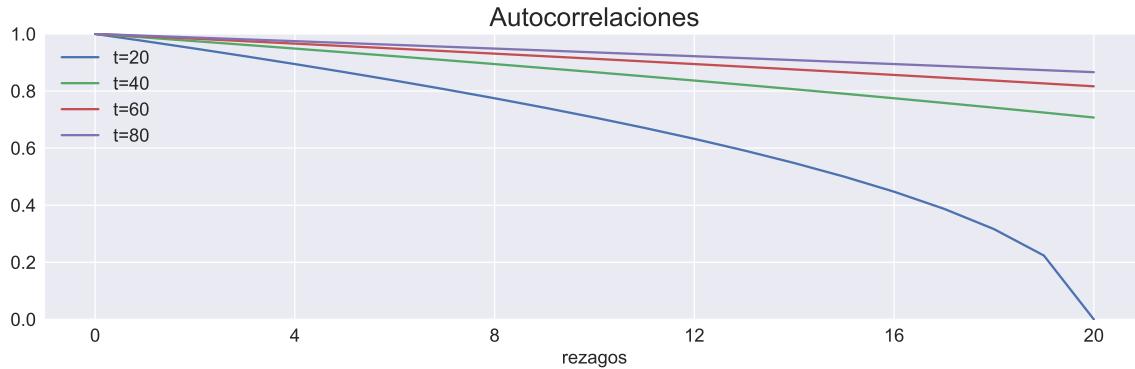
$$\Delta y_t = a + \epsilon_t \quad \text{o bien} \quad y_t = a + y_{t-1} + \epsilon_t$$

- Integrando el proceso sabemos que

$$y_t = y_0 + at + \sum_{\tau=1}^t \epsilon_\tau$$

- Vemos que la caminata aleatoria con deriva le añade una tendencia lineal determinística at a la caminata aleatoria convencional $y_0 + \sum_{\tau=1}^s \epsilon_{t+\tau}$
- Este es un ejemplo de un modelo de tendencia puro: no tiene un componente estacionario.
- Sabemos que añadir un componente determinístico a una variable aleatoria cambia su media en la misma magnitud de ese componente, pero no altera su varianza.
- Entonces, para la caminata aleatoria con deriva:

$$\begin{aligned}\mathbb{E} y_t &= y_0 + at & \text{Cov}(y_t, y_{t-s}) &= (t-s)\sigma^2 \\ \text{Var } y_t &= t\sigma^2 & \rho_{t,s} &= \sqrt{1 - \frac{s}{t}}\end{aligned}$$



Pronosticando una caminata aleatoria

- Es sumamente difícil pronosticar una caminata aleatoria.
- Supongamos que tenemos datos hasta t y deseamos pronostica el valor de la serie en $t+k$.
- En este caso:

$$\mathbb{E}[y_{t+k} | y_t] = y_t \quad \text{Var}[y_{t+k} | y_t] = k\sigma^2$$

- Es decir, el mejor pronóstico para cualquier valor futuro de la serie es su valor observado más reciente.
- Pero la varianza de este pronóstico crece linealmente, resultando infinitamente grande conforme $k \rightarrow \infty$.

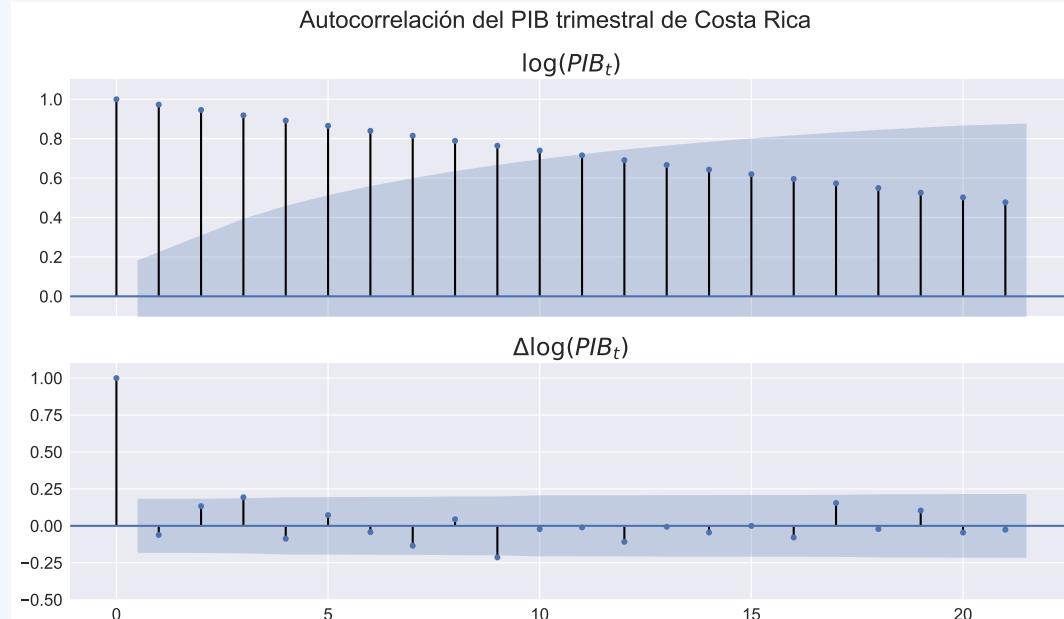
Series estacionarias en diferencia (DS)

- Las caminatas aleatorias (con o sin deriva) son ejemplos de series *estacionarias en diferencia*.
- Esto quiere decir que si tomamos su primera diferencia, el resultado es una serie estacionaria:

$$\begin{aligned}y_t &= y_0 + at + \epsilon_1 + \cdots + \epsilon_{t-1} + \epsilon_t \\ y_{t-1} &= y_0 + a(t-1) + \epsilon_1 + \cdots + \epsilon_{t-1} \\ \Rightarrow \Delta y_t &= a + \epsilon_t\end{aligned}$$

- Por ello, decimos que la serie y_t es *integrada de orden 1*.
- En general, si una serie z_t debe ser diferenciada d veces para obtener una serie estacionaria, entonces decimos que z_t es *integrada de orden d*, escrito $I(d)$.

Ejemplo 5-1: Correlación del PIB



5.3 Removiendo una tendencia

Distinguiendo una serie DS de una TS

- Comparemos la media y la varianza de estas dos series:

	TS	DS
Proceso	$x_t = x_0 + at + \epsilon_t$	$y_t = y_0 + at + \sum_{\tau=1}^t \epsilon_{\tau}$
Media	$x_0 + at$	$y_0 + at$
Varianza	σ^2	$t\sigma^2$
Δ	$a + \epsilon_t - \epsilon_{t-1}$	$a + \epsilon_t$
$-(\hat{z}_0 + \hat{a}t)$	ϵ_t	$\sum_{\tau=1}^t \epsilon_{\tau}$

- Los dos modelos tiene una media que crece linealmente, por lo que es muy difícil distinguirlos.

Ejemplo 5-2: Serie TS vs caminata aleatoria con deriva

```

import numpy as np
np.random.seed(1)
T = 121
e = np.random.randn(T)
e[0] = 0

x0 = y0 = a = 1
t = np.arange(T)

y = y0 + a*t + e.cumsum()
x = x0 + a*t + e

```

Este código simula una realización de serie estacionaria alrededor de tendencia (TS) y una estacionaria en diferencia (DS).

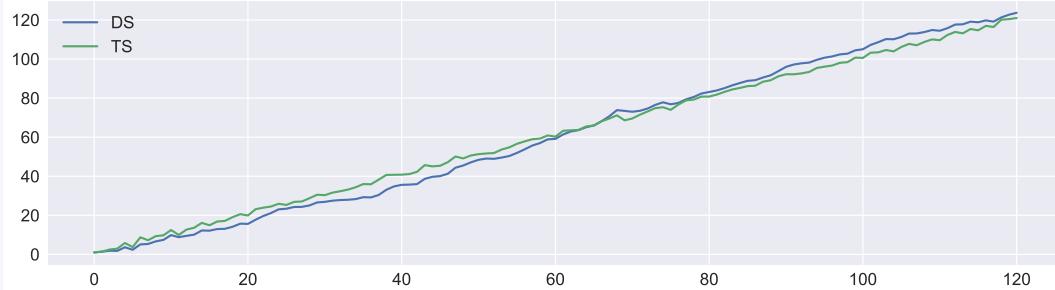
$$x_t = x_0 + at + \epsilon_t \quad (\text{TS})$$

$$y_t = y_0 + at + \sum_{\tau=1}^t \epsilon_{\tau} \quad (\text{DS})$$

Para que sea replicable, fijamos un valor para la función `np.random.seed`

Series no estacionarias

Tendencia determinística (TS) versus estocástica (DS)



- Hemos visto que para transformar en estacionaria una serie...
 - TS estimamos su tendencia determinística por regresión, usamos los residuos de esta regresión como el componente estacionario
 - DS tomamos su primera diferencia.
- Hecho esto, podemos modelar el componente estacionario, por ejemplo con ARMA.
- Pero es importante usar la técnica adecuada para remover la tendencia, según sea el tipo de serie (TS o DS)
- ¿Qué pasaría si tomamos primera diferencia de una serie TS?, o ¿qué pasaría si removemos una tendencia determinística de una serie DS?

Diferenciando una serie TS

(buscando problemas, parte 1)

- Para un proceso TS como $x_t = x_0 + at + \epsilon_t$, su primera diferencia es

$$\begin{aligned}\Delta x_t &= a + \epsilon_t - \epsilon_{t-1} \\ &= a + (1 - L)\epsilon_t\end{aligned}$$

- Es decir, la serie resultante Δx_t es un proceso MA(1).
- El polinomio de rezagos para la parte MA tiene una raíz unitaria, por lo que el proceso *no es invertible*.
- Esto puede causar problemas a la hora de estimar el modelo.

- Este problema también se presenta si diferenciamos una serie que ya es estacionaria: resultamos con un proceso no invertible y que tiene mayor varianza que el proceso original. A esto lo llamamos *sobrediferenciar*.

Restando una tendencia determinística a una serie DS

(buscando problemas, parte 2)

- Para un proceso DS como $y_t = y_0 + at + \sum_{\tau=1}^t \epsilon_{\tau}$, remover la tendencia determinística resulta en

$$\tilde{y}_t = \sum_{\tau=1}^t \epsilon_{\tau}$$

- Es decir, la serie resultante \tilde{y}_t aún conserva su tendencia estocástica.
- Vemos que

$$\begin{aligned}\Delta \tilde{y}_t &= \epsilon \\ \tilde{y}_t &= 1\tilde{y}_{t-1} + \epsilon\end{aligned}$$

- Esto es un proceso AR(1) con coeficiente autorregresivo igual a uno, por lo que no es estacionario.

5.4 Series con raíz unitaria y procesos ARIMA

Series con raíz unitaria

- La caminata aleatoria (con o sin deriva) puede verse como un modelo AR(1) en el cual la raíz del polinomio de rezagos es uno:

$$\begin{aligned}y_t &= c + 1y_{t-1} + \epsilon_t \\ (1 - 1L)y_t &= c + \epsilon_t\end{aligned}$$

- Por ello, en general decimos que un proceso con tendencia estocástica tiene raíz unitaria.
- Consideremos el siguiente proceso AR(2)

$$y_t = 0.6y_{t-1} + 0.4y_{t-2} + \epsilon_t$$

- Su polinomio de rezagos es $1 - 0.6L - 0.4L^2 = (1 - 0.4L)(1 - 1L)$, cuyas raíces son 1 y 2.5. Es decir, este proceso también tiene una raíz unitaria, y por tanto no es estacionario.
- La factorización del polinomio de rezagos nos sugiere escribir

$$\begin{aligned}(1 - 0.4L)(1 - 1L)y_t &= \epsilon_t \\ (1 - 0.4L)\Delta y_t &= \epsilon_t \\ \Delta y_t &= 0.4\Delta y_{t-1} + \epsilon_t\end{aligned}$$

- Visto como un proceso para $y^* \equiv \Delta y_t$, este es un proceso AR(1) estacionario.

Procesos ARIMA(p,d,q)

- En el ejemplo anterior, como la primera diferencia de y_t es un proceso AR(1) estacionario, decimos que y_t es un proceso ARIMA(1,1,0).
- En general, si un proceso integrado de orden d , y_t , es diferenciado d veces y su resultado $\Delta^d y_t$ es un proceso ARMA(p,q), entonces decimos que y_t tiene un proceso ARIMA(p,d,q):

Autorregresivo integrado de media móvil: ARIMA

Sea $\{\epsilon_t\}$ ruido blanco; el proceso estocástico

$$\Phi(L)(1 - L)^d y_t = \Theta(L)\epsilon_t$$

es llamado proceso ARIMA(p,d,q), donde $\Phi(L)$ es un polinomio de grado p cuyas raíces están fuera del círculo unitario, y $\Theta(L)$ es un polinomio de grado q .

Nota: Repaso del modelo clásico de regresión lineal

- En el modelo clásico de regresión lineal se tiene

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + \epsilon_i$$

donde $\epsilon_i \sim N(0, \sigma^2)$ es un error homoscedástico y no-autocorrelacionado.

- Si se cumplen los supuestos del MCRL, este modelo se puede estimar de manera insesgada y eficiente por medio del estimador mínimos cuadrados ordinarios.
- Para hacer un test sobre un parámetro

$$H_0 : \beta_j = q \quad H_1 : \beta_j \neq q$$

utilizamos el valor estimado por mínimos cuadrados ordinarios $\hat{\beta}_j$ y su error estándar $s.e.(\beta_j)$, y decimos que *si la hipótesis nula es cierta* entonces el estadístico

$$\frac{\hat{\beta}_j - q}{s.e.(\beta_j)} \sim t_{n-k}$$

(tiene una distribución *t*-Student con $n - k$ grados de libertad, donde n es el número de observaciones y k el número de parámetros estimados).

- Es decir, con un nivel de significancia α , el intervalo

$$[q + t_{\alpha/2} s.e.(\beta_j), q + t_{1-\alpha/2} s.e.(\beta_j)]$$

contendrá al valor estimado $\hat{\beta}_j$ en $100(1 - \alpha)\%$ de las muestras *siempre y cuando sea cierto que $\beta_j = q$* .

- Por ello, cuando encontramos un valor $\hat{\beta}_j$ que no está contenido en ese intervalo, rechazamos la hipótesis nula, y decimos que β_j es significativamente distinto de q , *a sabiendas de que nuestro procedimiento incurrirá en el error tipo-1 (rechazar una hipótesis verdadera) en $100\alpha\%$ de las muestras*.
- Para que este procedimiento tenga validez, es necesario que el estadístico efectivamente tenga la distribución *t*-Student, lo cual es cierto *siempre que se cumplan los supuestos del modelo clásico de regresión lineal*.

Regresión espuria

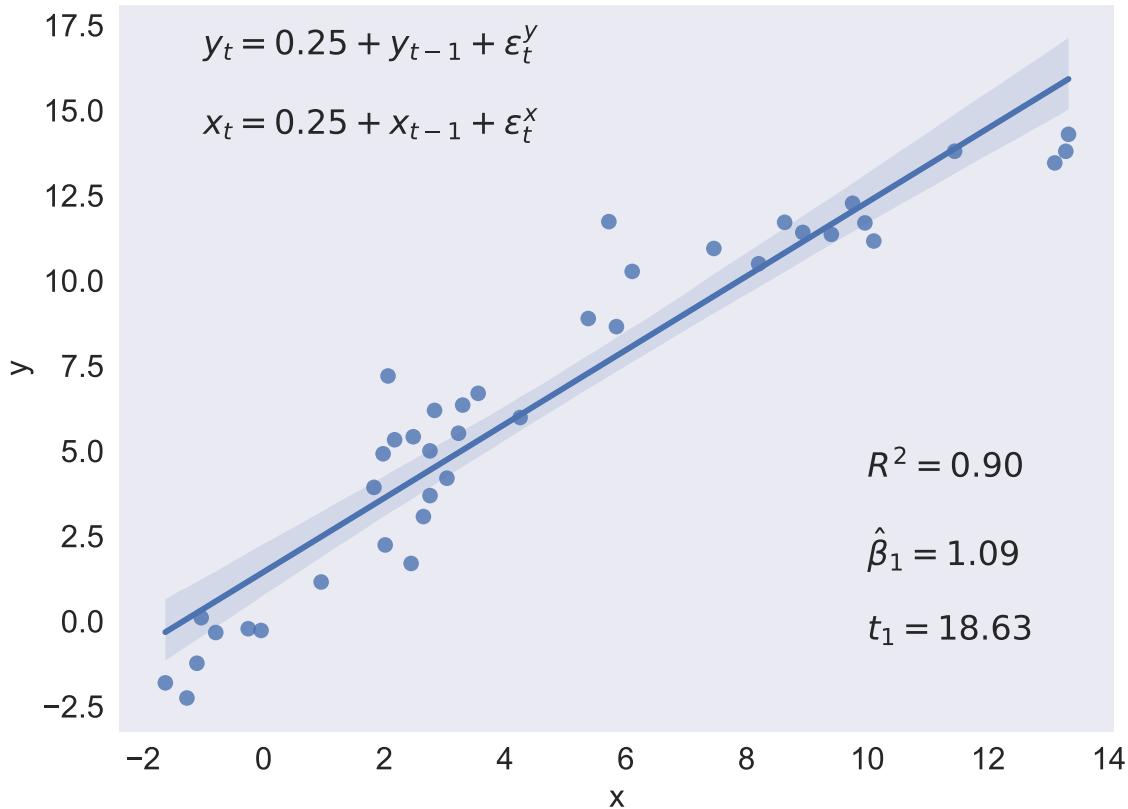
- En 1974, Granger y Newbold demostraron, via simulaciones, que si una serie I(1) se estima en función de otra serie I(1) completamente independiente de ella, los estadísticos usuales tenderán a mostar que las dos series están relacionadas.
- Es decir, si y_t, x_t son dos series I(1) independientes, al correr la regresión

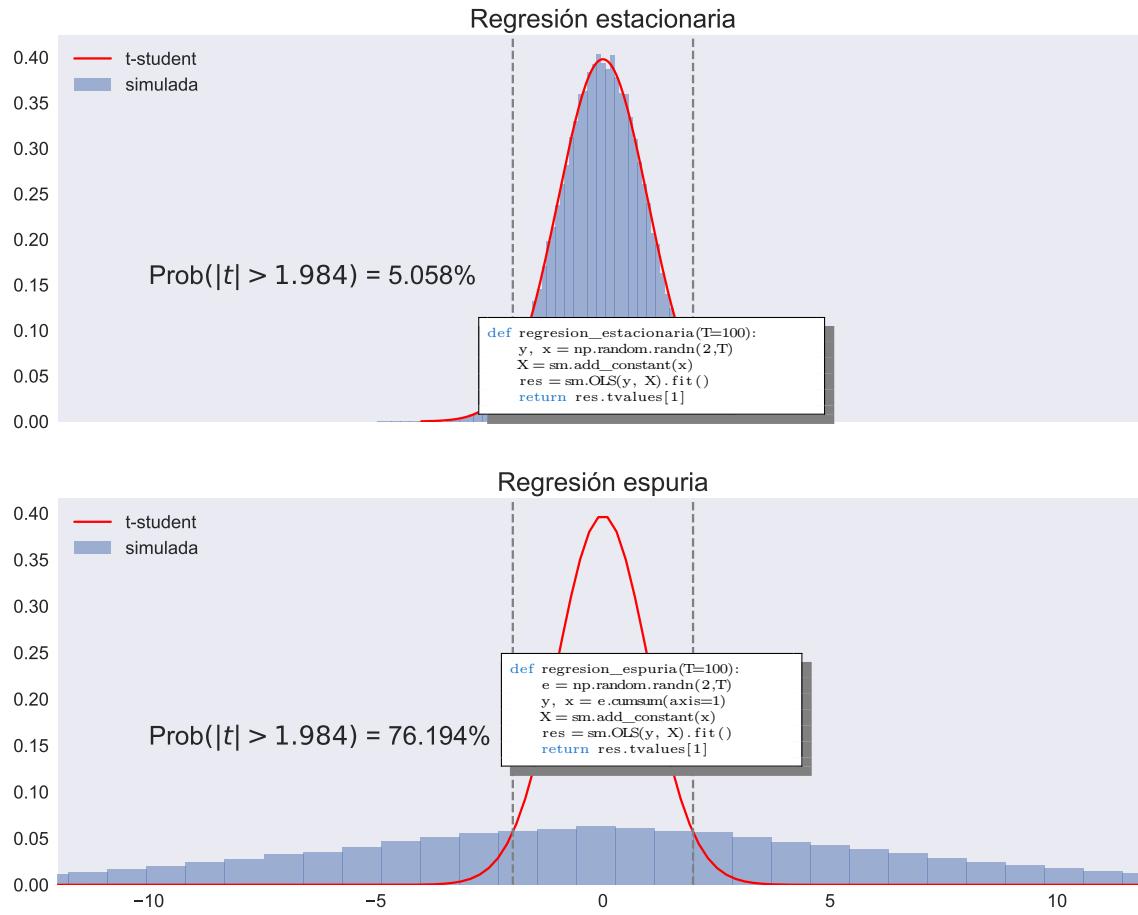
$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

fallaríamos en rechazar la hipótesis nula $\beta_1 = 0$ (la cual es cierta porque y_t no depende de x_t) con una frecuencia mayor a la que sugiere la distribución *t*-student correspondiente:

$$\frac{\hat{\beta}_1}{\hat{s.e}(\beta_1)} \sim t_{T-2} \Leftarrow \text{¡ya no es cierto!}$$

- Este fenómeno de encontrar relaciones inexistentes entre variables integradas se conoce como *regresión espuria*.





Regresión lineal con series integradas

¿Cómo estudiar la relación entre series integradas sin incurrir en regresiones espurias?

- Supongamos que tanto y_t como x_t son series I(1).
- Entonces, por definición, Δy_t y Δx_t son series estacionarias.
- En tal caso, el modelo

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \epsilon_t$$

puede estimarse por mínimos cuadrados, y la prueba $\beta_1 = 0$ se realiza usando el procedimiento usual.

- Una alternativa mejor, que estudiaremos con detalladamente más tarde en el curso, es estimar la regresión en niveles

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

y determinar si los residuos de esta regresión son estacionarios. Si lo son, diremos que y_t está cointegrada con x_t .

Procesos AR(p) con raíz unitaria

- Recordemos que el proceso AR(p) puede escribirse

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \\ (1 - \phi_1 L^1 - \cdots - \phi_p L^p) y_t &= c + \epsilon_t \\ \Phi(L) y_t &= c + \epsilon_t \end{aligned}$$

- Supongamos que los coeficientes autorregresivos suman uno.
- Entonces

$$\begin{aligned} \Phi(1) &= 1 - \phi_1 1^1 - \cdots - \phi_p 1^p \\ &= 1 - (\phi_1 + \phi_2 + \cdots + \phi_p) = 1 - 1 = 0 \end{aligned}$$

- Es decir, si $\phi_1 + \phi_2 + \cdots + \phi_p = 1$, entonces el proceso tiene raíz unitaria.

5.5 Detectando raíces unitarias

5.5.1 La prueba Dickey-Fuller

Caminata aleatoria como serie AR(1)

- El modelo más sencillo de una serie con raíz unitaria, la caminata aleatoria, es un proceso AR(1)

$$y_t = \phi y_{t-1} + \epsilon_t$$

en el cual se cumple que $\phi = 1$.

- Entonces resulta natural, para determinar si una serie es una caminata aleatoria, estimar esta ecuación y comprobar la hipótesis $\phi = 1$.
- Alternativamente, restando y_{t-1} de ambos lados podemos estimar

$$\begin{aligned} y_t - y_{t-1} &= (\phi - 1)y_{t-1} + \epsilon_t \\ \Delta y_t &= \gamma y_{t-1} + \epsilon_t \end{aligned}$$

y comprobar si

$$H_0 : \gamma_0 = 0 \quad \text{versus} \quad H_1 : \gamma_0 < 0$$

- No obstante, Dickey y Fuller (1979) encontraron que si la hipótesis nula es verdadera, la regresión anterior tiene series no estacionarias en ambos lados de la ecuación, por lo que no se cumple que

$$z = \frac{\hat{\gamma}}{s.e.(\gamma)}$$

tenga una distribución *t*-Student.

- Para determinar la distribución de este estadístico, de manera que pueda realizarse la prueba de hipótesis, Dickey y Fuller realizaron experimentos de Monte Carlo, en los cuales
 - Se simula una caminata aleatoria con un tamaño de muestra predeterminado.
 - Se estima el modelo AR(1)
 - Se calcula el valor de z
- Realizando muchas simulaciones como la anterior es posible aproximar la verdadera distribución del estadístico z *bajo la hipótesis nula* $\gamma = 0$.

Las pruebas de Dickey-Fuller

Las pruebas de Dickey-Fuller

?

¿Tiene la serie y_t una raíz unitaria?

H₀

$$\left. \begin{array}{l} \Delta y_t = \gamma y_{t-1} + \epsilon_t \\ \Delta y_t = a_0 + \gamma y_{t-1} + \epsilon_t \\ \Delta y_t = a_0 + b_0 t + \gamma y_{t-1} + \epsilon_t \end{array} \right\} \gamma = 0$$

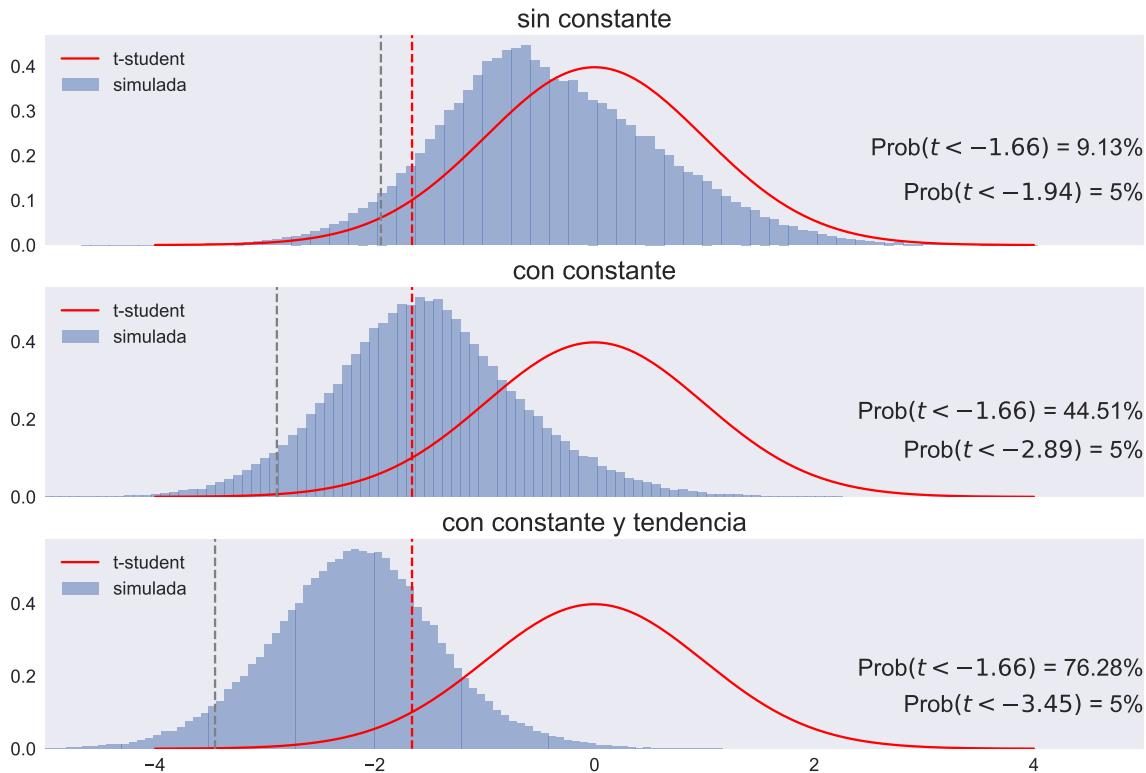
Test

Estimar $z = \frac{\hat{\gamma}}{s.e.(\gamma)}$ por mínimos cuadrados.

💡 Si z es menor que el valor crítico de Dickey Fuller, entonces $\gamma < 0$, es decir, no hay raíz unitaria.

La distribución de Dickey-Fuller

Distribución de Dickey-Fuller para T=100 observaciones



Ejemplo 5-3: Pruebas Dickey-Fuller

📄 bccr.ServicioWeb()(33783)

📄 tendencias-pib.ipynb

- Al estimar por mínimos cuadrados ordinarios la regresión

$$\Delta \log \text{PIB}_t = c + \phi \log \text{PIB}_{t-1} + \epsilon_t$$

encontramos

	coef	std err	t	P> t 	[0.025	0.975]
const	0.1010	0.042	2.393	0.018	0.017	0.185
IPIB	-0.0059	0.003	-2.143	0.034	-0.011	-0.000

- Los resultados de la tabla indican que ϕ es significativamente distinto de cero al 5% de significancia, *pero este resultado es incorrecto* porque en esta regresión el estadístico t no tiene la distribución t -Student.
- Además, la prueba reportada es de dos colas, mientras que la apropiada es de una cola.
- Por ello, recurrimos a los valores críticos de Dickey-Fuller

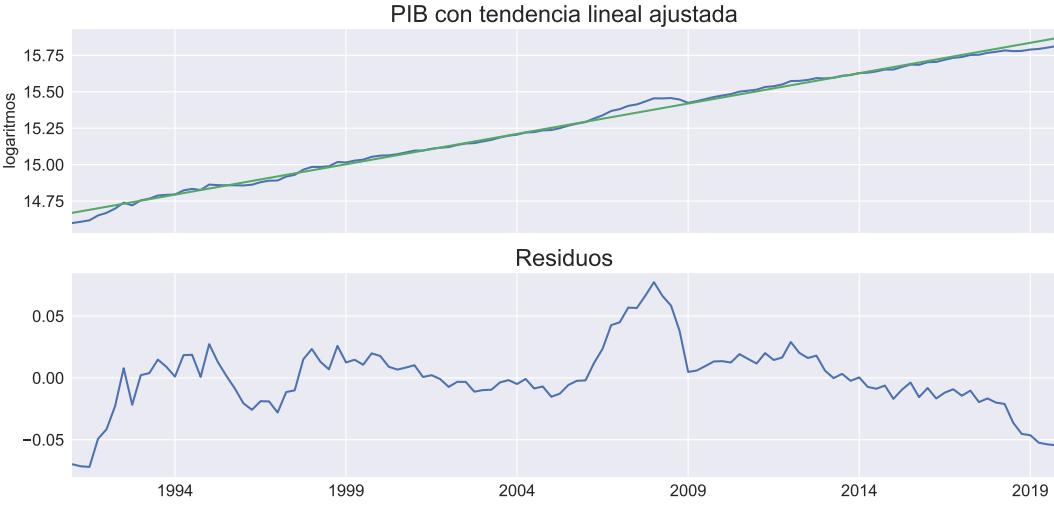
	z	1%	5%	10%
sin constante	10.709	-2.585	-1.944	-1.615
con constante	-2.143	-3.489	-2.887	-2.580
con constante y tendencia	-2.483	-4.040	-3.449	-3.150

- Como en todos los casos el valor z estimado es mayor que el valor crítico de Dickey-Fuller (sin importar cuál nivel de significancia utilizamos), no podemos rechazar la hipótesis de que el PIB tenga raíz unitaria.
- Por otra parte, si realizamos las pruebas de Dickey-Fuller a la primera diferencia del (logaritmo del) PIB trimestral encontramos

	z	1%	5%	10%
sin constante	-6.344	-2.586	-1.944	-1.615
con constante	-11.254	-3.489	-2.887	-2.580
con constante y tendencia	-11.641	-4.040	-3.450	-3.150

- Como en todos los casos el valor z estimado es menor que el valor crítico de Dickey-Fuller (sin importar cuál nivel de significancia utilizamos), concluimos que el crecimiento trimestral del PIB es estacionario (no tiene raíz unitaria).
- Dado que no pudimos rechazar que el PIB tuviese raíz unitaria, pero sí lo hicimos para su primer diferencia, concluimos que el PIB es una serie I(1).
- ¿Qué hubiera pasado si en vez de diferenciar la serie, le extraemos una tendencia lineal?
- Obtenemos los residuos de la regresión

$$\log \text{PIB}_t = c + at + \epsilon_t$$



Al aplicar las pruebas de Dickey-Fuller a los residuos, vemos que

- la prueba sin constante rechaza la presencia de una raíz unitaria, con 5% de significancia
- las otras dos pruebas no rechazan esta hipótesis.

	z	1%	5%	10%
sin constante	-2.523	-2.585	-1.944	-1.615
con constante	-2.515	-3.489	-2.887	-2.580
con constante y tendencia	-2.483	-4.040	-3.449	-3.150

5.5.2 La prueba aumentada de Dickey-Fuller

Caminata aleatoria como serie AR(p)

- No todas las series de tiempo pueden representarse apropiadamente como un proceso AR(1).
- La prueba de Dickey-Fuller puede aplicarse en estos casos también, aunque con modificaciones.
- Consideremos por ejemplo un proceso AR(2):

$$\begin{aligned}
 y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
 y_t &= (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2 y_{t-2} + y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
 y_t - y_{t-1} &= (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2(y_{t-1} - y_{t-2}) + \epsilon_t \\
 \Delta y_t &= (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2 \Delta y_{t-1} + \epsilon_t \\
 \Delta y_t &= \gamma y_{t-1} + a_1 \Delta y_{t-1} + \epsilon_t
 \end{aligned}$$

- Esta serie tiene raíz unitaria si $\gamma = 0$.

La prueba aumentada de Dickey-Fuller

- Para permitir la posibilidad que la serie original sea AR(p+1), la *prueba aumentada de Dickey-Fuller* introduce p rezagos de la variable dependiente en la regresión original:

$$\begin{aligned}\Delta y_t &= \gamma y_{t-1} + a_1 \Delta y_{t-1} + \cdots + a_p \Delta y_{t-p} + \epsilon_t \\ \Delta y_t &= a_0 + \gamma y_{t-1} + a_1 \Delta y_{t-1} + \cdots + a_p \Delta y_{t-p} + \epsilon_t \\ \Delta y_t &= a_0 + b_0 t + \gamma y_{t-1} + a_1 \Delta y_{t-1} + \cdots + a_p \Delta y_{t-p} + \epsilon_t\end{aligned}$$

- En cualquiera de las formulaciones, la hipótesis nula es $\gamma = 0$.
- Se utilizan los mismos valores críticos de la prueba de Dickey-Fuller.

Ejemplo 5-4: Pruebas aumentada de Dickey-Fuller

 bccr

 tendencias-pib.ipynb

- Al realizar las pruebas aumentadas de Dickey-Fuller del (logaritmo del) PIB trimestral de Costa Rica encontramos

	z	1%	5%	10%	p
sin constante	3.985	-2.587	-1.944	-1.615	9
con constante	-1.559	-3.494	-2.889	-2.582	9
con constante y tendencia	-1.586	-4.047	-3.453	-3.152	9

- Esto confirma lo que ya habíamos encontrado: no podemos rechazar la hipótesis de que el PIB tenga raíz unitaria.
- En todos los casos, el número de rezagos corresponde al máximo rezago significativo.
- Por otra parte, si realizamos las pruebas de Dickey-Fuller a la primera diferencia del (logaritmo del) PIB trimestral encontramos

	z	1%	5%	10%	p
sin constante	-1.694	-2.587	-1.944	-1.615	7
con constante	-4.465	-3.494	-2.889	-2.582	8
con constante y tendencia	-4.719	-4.047	-3.453	-3.152	8

- De nuevo concluimos que el crecimiento trimestral del PIB es estacionario (no tiene raíz unitaria).

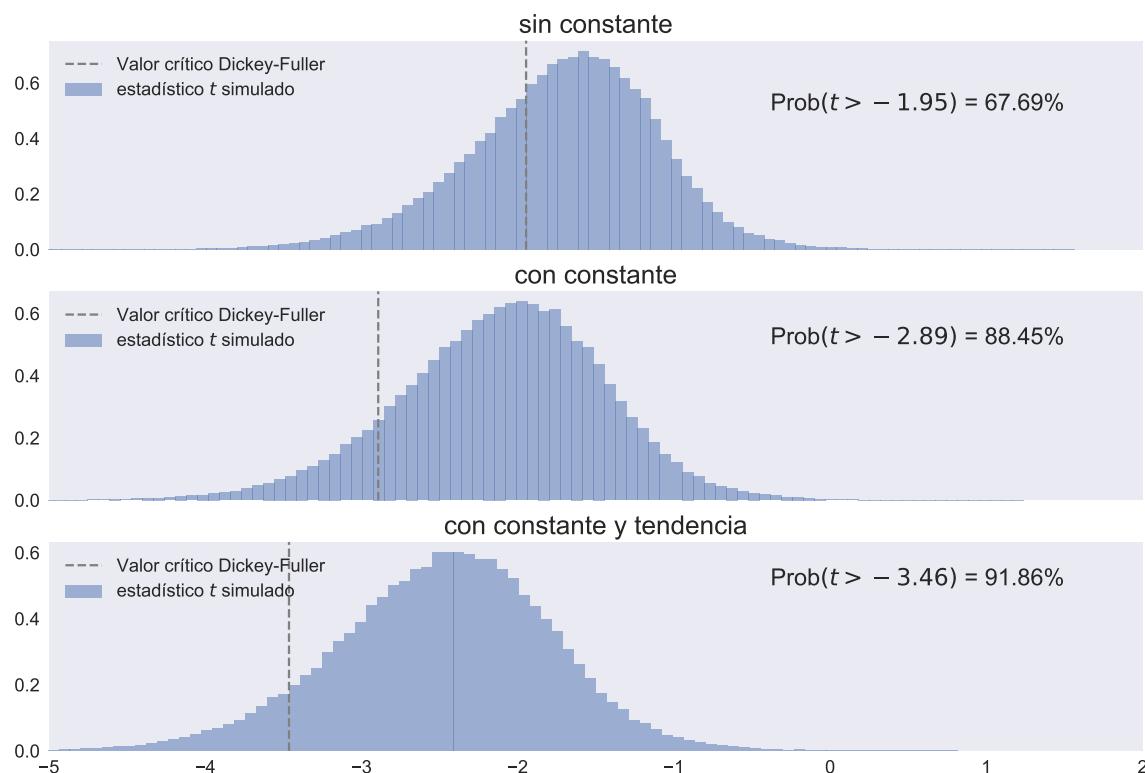
Interpretando una prueba de Dickey Fuller

- Importante:* En la prueba DF, no rechazar la hipótesis de que una serie tenga raíz unitaria...

- *no implica* que la serie sí tenga tal raíz unitaria.
- solamente decimos que *no hay evidencia* suficiente para descartarla con un nivel “razonable” de significancia.
- Esto es así porque bien podría ser el caso de que el verdadero valor de ϕ sea ligeramente menor a uno (en cuyo caso el proceso AR(1) es estacionario), pero la prueba Dickey-Fuller no puede distinguirlo efectivamente de 1.

Potencia de la distribución de Dickey-Fuller

Distribución de estadístico t para AR(1) cuando $\rho = 0.95$, con $T=100$ observaciones,



Limitaciones de las pruebas de raíz unitaria

- Hemos visto que la prueba de Dickey-Fuller tienen muy poca potencia para casos en que el proceso es persistente pero no integrado.
- Esto es una limitación también de otras pruebas de raíz unitaria, como la de Phillips-Perron.
- Por ello, cuando se estudian series macroeconómicas con estos tests, usualmente se encuentra que tienen raíces unitarias.
- Esto se debe a que la hipótesis nula es que sí hay raíz unitaria, y esta hipótesis solamente se rechaza cuando existe *fuerte evidencia* en su contra.

5.5.3 La prueba KPSS

La prueba KPSS

- Kwiatkowski, Phillips, Schmidt y Shin (1992) proponen una *prueba de estacionariedad*: la hipótesis nula es que la serie es estacionaria.
- Para ello, asumen que una serie puede ser expresada como la suma de una tendencia determinística, una caminata aleatoria, y un error estacionario (no necesariamente ruido blanco):

$$y_t = \underset{\text{tendencia}}{\xi t} + \underset{\text{caminata}}{r_t} + \underset{\text{estacionario}}{\omega_t}$$

donde r_t es una caminata aleatoria

$$r_t = r_{t-1} + u_t, \quad u_t \sim N(0, \sigma_u^2)$$

- La hipótesis de estacionariedad es simplemente $\sigma_u^2 = 0$.
- Bajo la hipótesis nula, $r_t = r_{t-1} = \dots = r_0$ una constante, por lo que la serie sería estacionaria alrededor de una tendencia:

$$y_t = r_0 + \xi t + \omega_t$$

- KPSS también consideran el caso particular en el que $\xi = 0$, es decir, la serie es simplemente estacionaria.

$$y_t = r_0 + \omega_t$$

- En cualquiera de estos dos casos, si e_1, e_2, \dots, e_T son los residuos de la regresión, se define

$$\begin{aligned} S_t &= \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, T && \text{(suma parcial de residuos)} \\ \hat{\sigma}_e^2 &= \frac{1}{T} \sum_{t=1}^T e_t^2 && \text{(varianza estimada del error)} \end{aligned}$$

- $\hat{\sigma}_e^2$ es un estimador consistente de la varianza de la parte estacionaria ω_t solo si es ruido blanco.
- Pero en la práctica, las series económicas rara vez cumplen esa restricción, por lo que KPSS proponen esta corrección para tomar en cuenta la posible correlación de ω_t :

$$s^2(l) = \frac{1}{T} \sum_{t=1}^T e_t^2 + \frac{2}{T} \sum_{s=1}^l \left[\left(1 - \frac{s}{1+l} \right) \sum_{t=s+1}^T e_t e_{t-s} \right]$$

- Así, para hacer una prueba KPSS hay que decidir:

- si incluir o no la tendencia determinística
- cuántos rezagos l incluir en la estimación de la varianza $s^2(l)$

La prueba KPSS



¿Es la serie y_t estacionaria?

H₀

$$y_t = \xi t + r_t + \omega_t$$

$$r_t = r_{t-1} + u_t, \quad u_t \sim N(0, \sigma_u^2)$$

$$\sigma_u^2 = 0$$

Test

$$LM = \sum_{t=1}^T \frac{S_t^2}{s^2(l)}$$



Si LM es mayor que el valor crítico, rechazar la hipótesis nula y concluir que la serie tiene raíz unitaria.

- KPSS proporcionan los siguientes valores críticos asintóticos, los cuales obtuvieron por simulación 50 000 iteraciones con muestras de 2000 datos.

	c	ct
10%	0.347	0.119
5%	0.463	0.146
2.5%	0.574	0.176
1%	0.739	0.216

- Las pruebas son de una cola: se rechaza la hipótesis nula (de que la serie es estacionaria) cuando el estadístico LM es mayor al valor crítico seleccionado.

Ejemplo 5-5: Pruebas KPSS

bccr

tendencias-pib.ipynb

- Al realizar las pruebas KPSS del (logaritmo del) PIB trimestral de Costa Rica encontramos

nivel	diferencia			
	c	ct	c	ct
0	11.569	1.186	0.371	0.046
1	5.864	0.632	0.395	0.051
2	3.952	0.444	0.368	0.049
3	2.994	0.350	0.326	0.044
4	2.419	0.296	0.313	0.043
5	2.035	0.262	0.300	0.042
6	1.762	0.238	0.294	0.042

Valores críticos

	c	ct
10%	0.347	0.119
5%	0.463	0.146
2.5%	0.574	0.176
1%	0.739	0.216

- En todas las especificaciones, podemos rechazar al 1% que el PIB sea estacionario.
- Al 5%, en ningún caso podemos rechazar que el crecimiento del PIB (diferencia del logaritmo) sea estacionaria.
- Juntos, estos resultados señalan que el PIB es un proceso I(1).

Ejemplo 5-6: Raíces unitarias en series macroeconómicas de Estados Unidos

NelsonPlosser.ipynb

- En uno de los artículos más citados en macroeconomía, Nelson y Plosser (1982) examinaron varias series macro de uso común, averiguando si tenían raíces unitarias.
- Aplicando la prueba aumentada de Dickey-Fuller, concluyeron que todas menos una de las series analizadas tenían raíz unitaria.

Table 5
Tests for autoregressive unit roots*

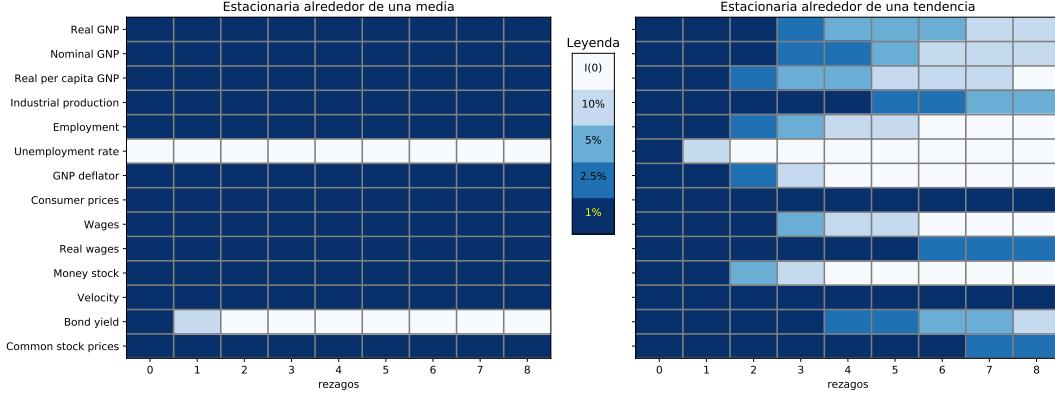
$$z_t = \hat{\mu} + \hat{\gamma}t + \hat{\rho}_1 z_{t-1} + \hat{\rho}_2(z_{t-1} - z_{t-2}) + \cdots + \hat{\rho}_k(z_{t-k+1} - z_{t-k}) + \hat{\epsilon}_t$$

Series	T	k	$\hat{\mu}$	t($\hat{\mu}$)	$\hat{\gamma}$	t($\hat{\gamma}$)	$\hat{\rho}_1$	t($\hat{\rho}_1$)	s($\hat{\mu}$)	r _t
Real GNP	62	2	0.819	3.03	0.006	3.03	0.825	-2.99	0.058	-0.02
Nominal GNP	62	2	1.06	2.37	0.006	2.34	0.899	-2.32	0.087	0.03
Real per capita GNP	62	2	1.28	3.05	0.004	3.01	0.818	-3.04	0.059	-0.02
Industrial production	111	6	0.103	4.32	0.007	2.44	0.835	-2.53	0.097	0.03
Employment	81	3	1.42	2.68	0.002	2.54	0.861	-2.66	0.035	0.10
Unemployment rate	81	4	0.513	2.81	-0.000	-0.23	0.706	-3.55*	0.407	0.02
GNP deflator	82	2	0.260	2.55	0.002	2.65	0.915	-2.52	0.046	-0.03
Consumer prices	111	4	0.090	1.76	0.001	2.84	0.986	-1.97	0.042	-0.06
Wages	71	3	0.566	2.30	0.004	2.30	0.910	-2.09	0.060	0.00
Real wages	71	2	0.487	3.10	0.004	3.14	0.831	-3.04	0.034	-0.01
Money stock	82	2	0.133	3.52	0.005	3.03	0.916	-3.08	0.047	0.03
Velocity	102	1	0.052	0.99	-0.000	-0.65	0.941	-1.66	0.067	0.11
Interest rate	71	3	-0.186	-0.95	0.003	1.75	1.03	0.686	0.283	-0.02
Common stock prices	100	3	0.481	2.02	0.003	2.37	0.913	-2.05	0.158	0.20

- En su artículo original, KPSS aplican su prueba a las mismas series que utilizaron Nelson y Plosser.
- Encontraron que para varias de las series no era posible rechazar la hipótesis de que fueran estacionarias alrededor de una tendencia.

Table 5
 $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ tests for trend stationarity applied to Nelson–Plosser data.

Series	Lag truncation parameter (l)								
	0	1	2	3	4	5	6	7	8
η_μ : 5% critical value is 0.463									
Real GNP	5.96	3.06	2.08	1.59	1.30	1.11	0.97	0.86	0.78
Nominal GNP	5.81	2.98	2.04	1.56	1.28	1.09	0.95	0.85	0.77
Real per capital GNP	5.54	2.84	1.94	1.50	1.22	1.05	0.92	0.82	0.75
Industrial production	10.79	5.48	3.70	2.81	2.27	1.92	1.66	1.47	1.32
Employment	7.57	3.87	2.63	2.01	1.64	1.39	1.21	1.08	0.98
Unemployment rate	0.31	0.18	0.14	0.11	0.10	0.10	0.09	0.09	0.09
GNP deflator	7.51	3.82	2.59	1.97	1.60	1.35	1.18	1.04	0.94
Consumer prices	7.90	4.02	2.73	2.08	1.69	1.43	1.24	1.10	0.99
Wages	6.72	3.43	2.33	1.78	1.45	1.23	1.07	0.95	0.86
Real wages	6.96	3.55	2.40	1.83	1.48	1.26	1.09	0.97	0.88
Money	8.01	4.08	2.76	2.10	1.70	1.44	1.25	1.11	1.00
Velocity	8.40	4.29	2.90	2.21	1.80	1.52	1.32	1.17	1.05
Interest rate	0.78	0.42	0.30	0.24	0.20	0.17	0.16	0.14	0.13
Stock prices	8.01	4.10	2.79	2.13	1.74	1.48	1.29	1.15	1.04
$\hat{\eta}_\tau$: 5% critical value is 0.146									
Real GNP	0.630	0.337	0.242	0.198	0.173	0.158	0.148	0.141	0.137
Nominal GNP	0.755	0.392	0.273	0.215	0.181	0.159	0.143	0.132	0.124
Real per capital GNP	0.528	0.283	0.204	0.167	0.147	0.134	0.126	0.121	0.118
Industrial production	0.822	0.446	0.320	0.257	0.220	0.196	0.179	0.166	0.155
Employment	0.526	0.278	0.198	0.158	0.136	0.122	0.112	0.105	0.101
Unemployment rate	0.216	0.124	0.094	0.079	0.071	0.066	0.063	0.061	0.061
GNP deflator	0.492	0.256	0.178	0.140	0.117	0.103	0.093	0.086	0.081
Consumer prices	1.85	0.943	0.641	0.491	0.401	0.342	0.301	0.270	0.246
Wages	0.612	0.317	0.220	0.173	0.145	0.128	0.115	0.107	0.101
Real wages	0.956	0.511	0.365	0.293	0.252	0.226	0.208	0.194	0.184
Money	0.445	0.228	0.158	0.124	0.104	0.092	0.084	0.079	0.075
Velocity	1.78	0.932	0.647	0.504	0.418	0.360	0.319	0.287	0.262
Interest rate	0.845	0.457	0.323	0.255	0.214	0.186	0.166	0.151	0.140
Stock prices	1.23	0.646	0.454	0.359	0.302	0.264	0.237	0.216	0.199



5.6 Otro método para remover una tendencia

Desagregación de una serie de tiempo

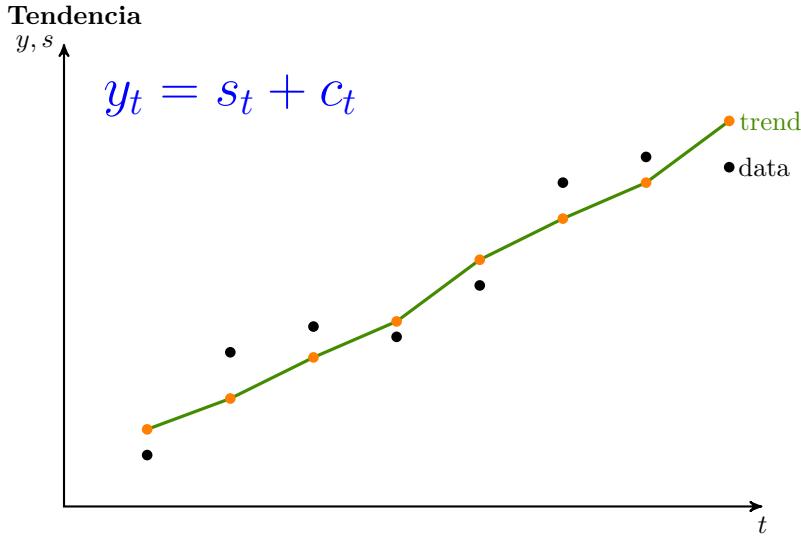
- Tenemos una muestra de T observaciones de la variable aleatoria Y_t :

$$\{y_1, y_2, \dots, y_T\}$$

- Y_t tiene dos componentes: crecimiento (tendencia) s_t y ciclo c_t .

$$y_t = s_t + c_t$$

- Asumimos que la tendencia es una curva suave, aunque no necesariamente una línea recta.



Objetivos en conflicto

- Partiendo de y_t , Hodrick y Prescott 1997 “extraen” la tendencia s_t

$$\{s_1, s_2, \dots, s_T\},$$

tratando de balancear dos objetivos mutuamente excluyentes:

1. el ajuste a los datos originales, es decir, $y_t - s_t$ debe ser pequeño.
 2. la tendencia resultante debe ser suave, por lo que los cambios de pendiente $(s_{t+1} - s_t) - (s_t - s_{t-1})$ también deben ser pequeños.
- La importancia relativa de estos dos factores es ponderada con el parámetro λ .

El filtro de Hodrick y Prescott

Formalmente, la tendencia la definen por:

$$s_i^{HP} = \underset{s_1, \dots, s_T}{\operatorname{argmin}} \left\{ \sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} [(s_{t+1} - s_t) - (s_t - s_{t-1})]^2 \right\}$$

$$= \underset{s_1, \dots, s_T}{\operatorname{argmin}} \left\{ \sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} (s_{t+1} - 2s_t + s_{t-1})^2 \right\}$$

Definimos las matrices

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_T \end{bmatrix}$$

$$A_{T-2 \times T} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ & & & & & \ddots & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

Reescribimos el problema de optimización

$$\begin{aligned} s_i^{HP} &= \underset{s_1, \dots, s_T}{\operatorname{argmin}} \left\{ \sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} (s_{t+1} - 2s_t + s_{t-1})^2 \right\} \\ &= \underset{S}{\operatorname{argmin}} \{(Y - S)'(Y - S) + \lambda(AS)'(AS)\} \\ &= \underset{S}{\operatorname{argmin}} \{Y'Y - 2Y'S + S'(I + \lambda A'A)S\} \end{aligned}$$

Resolviendo el problema

- Taking the FOC

$$S^{HP} = \underset{S}{\operatorname{argmin}} \{Y'Y - 2Y'S + S'(I + \lambda A'A)S\}$$

$$\Rightarrow -2Y + 2(I + \lambda A'A)S = 0$$

- Then, the HP filter is

$$\begin{aligned} S^{HP} &= (I + \lambda A'A)^{-1} Y && \text{(tendencia)} \\ C^{HP} \equiv Y - S^{HP} &= \left[I - (I + \lambda A'A)^{-1} \right] Y && \text{(ciclo)} \end{aligned}$$

Ejemplo 5-7: El filtro HP

Asumimos que tenemos $T = 5$ datos $Y = [y_1, y_2, y_3, y_4, y_5]'$ y que $\lambda = 4$. Los datos de tendencia $S = [s_1, s_2, s_3, s_4, s_5]'$ están dados por:

$$\begin{array}{lllll} s_1 = & 0.67y_1 & +0.36y_2 & +0.13y_3 & -0.02y_4 & -0.14y_5 \\ s_2 = & 0.36y_1 & +0.34y_2 & +0.23y_3 & +0.10y_4 & -0.02y_5 \\ s_3 = & 0.13y_1 & +0.23y_2 & +0.29y_3 & +0.23y_4 & +0.13y_5 \\ s_4 = & -0.02y_1 & +0.10y_2 & +0.23y_3 & +0.34y_4 & +0.36y_5 \\ s_5 = & -0.14y_1 & -0.02y_2 & +0.13y_3 & +0.36y_4 & +0.67y_5 \end{array}$$

Observe que cada dato de tendencia s_t es simplemente un promedio ponderado de todos los

datos en Y . Además, algunas de las ponderaciones son negativas! Por otra parte, los datos del ciclo $C = [c_1, c_2, c_3, c_4, c_5]'$ están dados por:

$$\begin{array}{llllll} c_1 = & 0.33y_1 & -0.36y_2 & -0.13y_3 & +0.02y_4 & +0.14y_5 \\ c_2 = & -0.36y_1 & +0.66y_2 & -0.23y_3 & -0.10y_4 & +0.02y_5 \\ c_3 = & -0.13y_1 & -0.23y_2 & +0.71y_3 & -0.23y_4 & -0.13y_5 \\ c_4 = & 0.02y_1 & -0.10y_2 & -0.23y_3 & +0.66y_4 & -0.36y_5 \\ c_5 = & 0.14y_1 & +0.02y_2 & -0.13y_3 & -0.36y_4 & +0.33y_5 \end{array}$$

De nuevo, observe que cada dato del ciclo c_t es un promedio ponderado de todos los puntos en Y , pero donde las ponderaciones suman cero.

Escogiendo λ

- El resultado del filtro es muy sensible a la escogencia de λ .
- Como regla habitual, λ se escoge según la frecuencia de los datos
 - Anuales $\Rightarrow 100$
 - Trimestrales $\Rightarrow 1600$
 - Mensuales $\Rightarrow 14400$

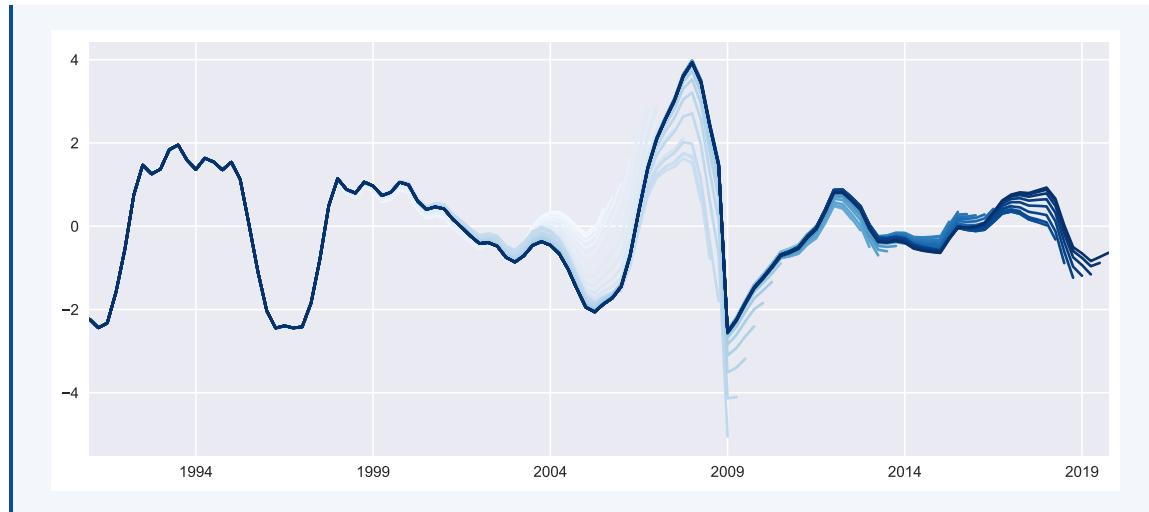
El filtro HP tiene malas propiedades estadísticas

Hamilton 2017: Why You Should Never Use the HP Filter?

1. El filtro Hodrick-Prescott introduce relaciones dinámicas espurias que no tienen sustento en el proceso generador de datos subyacente.
2. Los valores filtrados al final de la muestra son muy distintos de los del medio, y también están caracterizados por una dinámica espuria.
3. Una formalización estadística del problema típicamente produce valores de λ que distan mucho de los usados comúnmente en la práctica.
4. Para Hamilton, hay una alternativa mejor: una regresión AR(4) alcanza todos los objetivos buscados por usuarios del filtro HP pero con ninguno de sus desventajas.

Ejemplo 5-8: Filtrando el PIB de Costa Rica con HP

Los datos filtrados son muy sensibles a nueva información.



HP puede inducir conclusiones equivocadas acerca del comovimiento de series

Cogley y Nason (1995) analizaron las propiedades espectrales del filtro HP

- Cuando se mide el componente cíclico de una serie de tiempo, ¿es buena idea usar el filtro HP?
- Depende de la serie original
 - Sí**, si es estacionaria alrededor de tendencia
 - No**, si es estacionaria en diferencia
- Este resultado tiene implicaciones importantes para modelos DSGE
 - Cuando se aplica el filtro HP a una serie integrada, el filtro introduce periodicidad y comovimiento en las frecuencias del ciclo económico, *aún si no estaban presentes* en los datos originales.

Referencias del capítulo 5

- Cogley, Timothy y James M. Nason (1995). “Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research”. En: *Journal of Economic Dynamics and Control* 19.1–2, págs. 253-278. ISSN: 0165-1889.
- Enders, Walter (2015). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Hamilton, James D. (2017). “Why You Should Never Use the Hodrick-Prescott Filter”. En: *Review of Economics and Statistics* 100.5.
- Hodrick, Robert J. y Edward C. Prescott (1997). “Postwar U.S. Business Cycles: An Empirical Investigation”. En: *Journal of Money, Credit and Banking* 29.1, págs. 1-16. ISSN: 00222879, 15384616.
- Levendis, John D. (2018). *Time Series Econometrics. Learning Through Replication*. Springer. ISBN: 978-3-319-98281-6.

Capítulo 6

Cambio estructural

En este capítulo

6.1	Introducción	113
6.2	La prueba de Chow	114
6.3	Cambio estructural y raíces unitarias: fecha conocida	117
6.4	Cambio estructural y raíces unitarias: fecha desconocida	122

6.1 Introducción

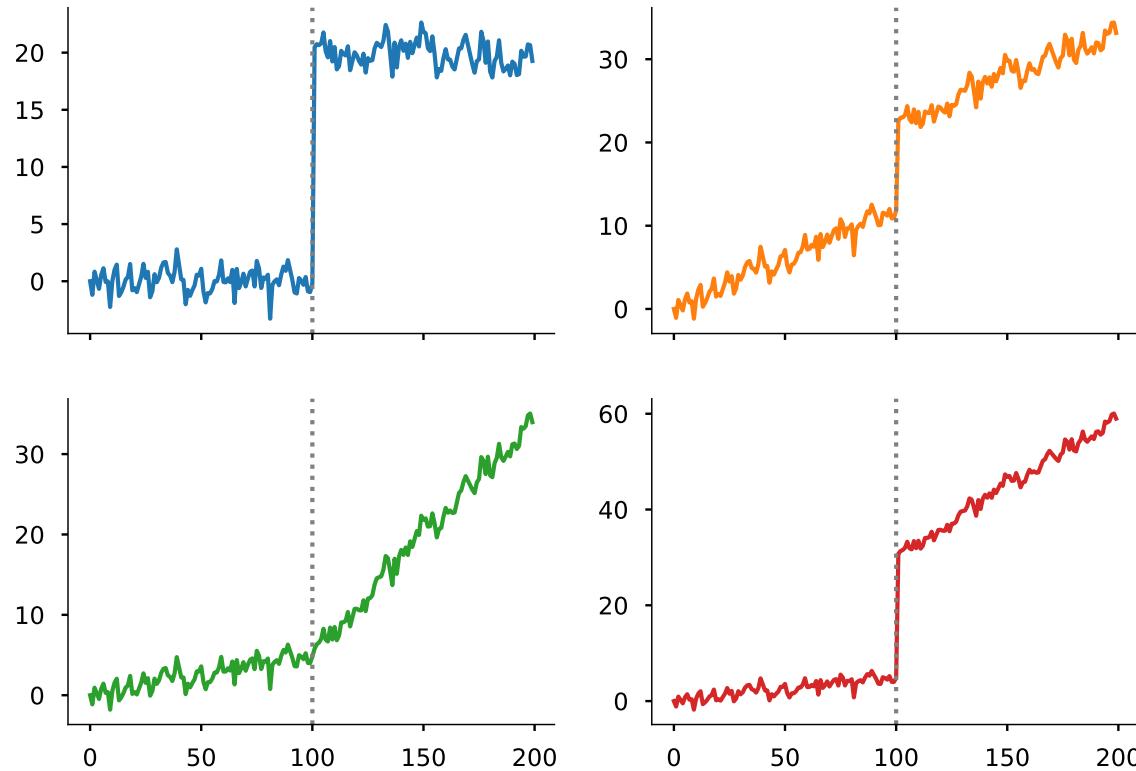
La crítica de Lucas

- En 1976 Lucas publicó una de las críticas más fuertes al enfoque de modelación económica de gran escala de la Comisión Cowles.
- Lucas criticó el supuesto de que muchos fenómenos económicos son estructurales.
- En esencia, argumentaba que cambios en las leyes y regulaciones afectaban el comportamiento humano, y que esto se reflejaría en los datos.
- Por ejemplo, en una estimación de consumo

$$c_t = \alpha + \beta y_t + \epsilon_t$$

la propensión marginal a consumir β podría cambiar con el tiempo.

Cambio estructural en una serie de tiempo



6.2 La prueba de Chow

Determinando si hay un cambio estructural

- Recordemos que podemos expresar una regresión lineal como

$$\begin{aligned} y_t &= \beta_1 x_{t,1} + \cdots + \beta_k x_{t,k} + \epsilon \\ &= [x_{t,1} \ \dots \ x_{t,k}] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \epsilon_t \\ &= x_t' \beta + \epsilon_t \end{aligned}$$

- Cuando apilamos todas T las observaciones

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_1' \beta + \epsilon_1 \\ \vdots \\ x_T' \beta + \epsilon_T \end{bmatrix} = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} = X\beta + \epsilon$$

- La regresión $Y = X\beta + \epsilon$ asume que el vector de coeficientes β es el mismo para toda la muestra.
- Supongamos que pensamos que durante el período conocido hubo un cambio estructural en la economía, por lo que el vector de parámetros fue β_0 antes del cambio pero β_1 después.

Entonces:

$$\begin{cases} Y_1 = X_1\beta_1 + \epsilon_1, & \text{antes del cambio} \\ Y_2 = X_2\beta_2 + \epsilon_2, & \text{después} \end{cases}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

- Por lo que el estimador de mínimos cuadrados ordinarios

$$\widehat{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}} = \begin{bmatrix} X'_1 X_1 & 0 \\ 0 & X'_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} X'_1 Y_1 \\ X'_2 Y_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

- Esto nos dice que los parámetros β_1 y β_2 pueden estimarse por MCO separadamente.
- Planteamos la hipótesis nula de que no hay cambio estructural: $\beta_1 = \beta_2$. Entonces

$$\begin{cases} Y_1 = X_1\beta + \epsilon_1, & \text{antes del cambio} \\ Y_2 = X_2\beta + \epsilon_2, & \text{todo sigue igual} \end{cases}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

- Esto corresponde a la regresión con todas las observaciones.
- La hipótesis $\beta_1 = \beta_2$ puede comprobarse con un test de Wald.

La prueba de Chow

- Sea S la suma de los cuadrados de los residuos de la regresión restringida, y S_i la suma de los cuadrados de los residuos de la regresión de la muestra i .

Prueba de cambio estructural de Chow	
	¿Son los coeficientes β_1 y β_2 distintos?
H_0	$\beta_1 = \beta_2$
Test	$\frac{\frac{S - S_1 - S_2}{k}}{\frac{S_1 + S_2}{T-2k}} \sim F(k, T - 2k)$
	Si el estadístico es grande (mayor que el valor crítico), los modelos son distintos, por lo que sí hay cambio estructural.

Ejemplo 6-1: Cambio en la demanda por combustible

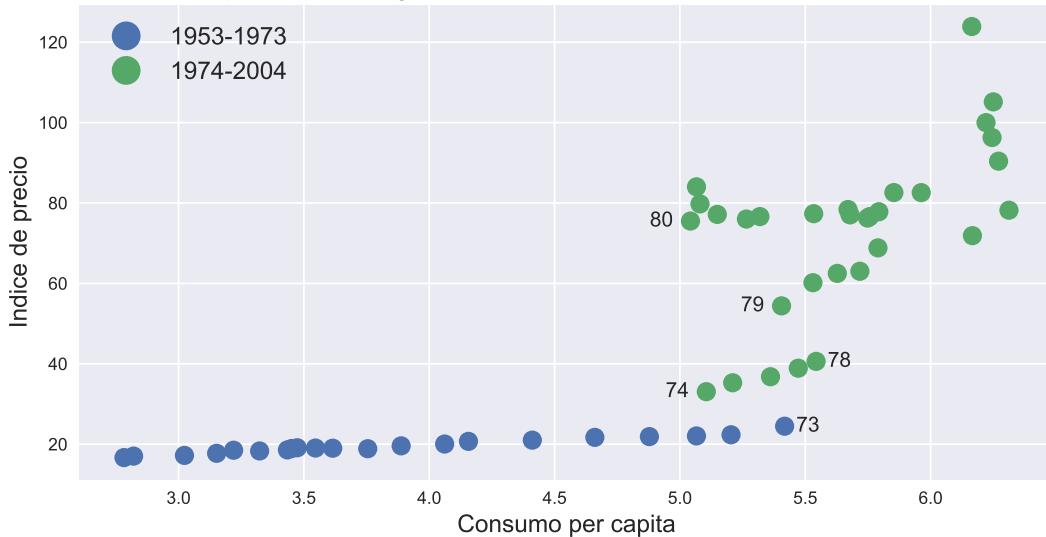


TableF2-2.txt



Greene-gas-market.ipynb

Consumo per capita de gasolina versus Precio, Estados Unidos 1953–2004



El mercado del petróleo tuvo un cambio significativo en 1973:



Antes de 1973, el comercio de petróleo estaba dominado por las “Siete Hermanas”, que controlaba $\approx 85\%$ de las reservas de petróleo



Desde la guerra de Yom Kippur (oct-1973), la OPEP ha dominado activamente en la fijación del precio.

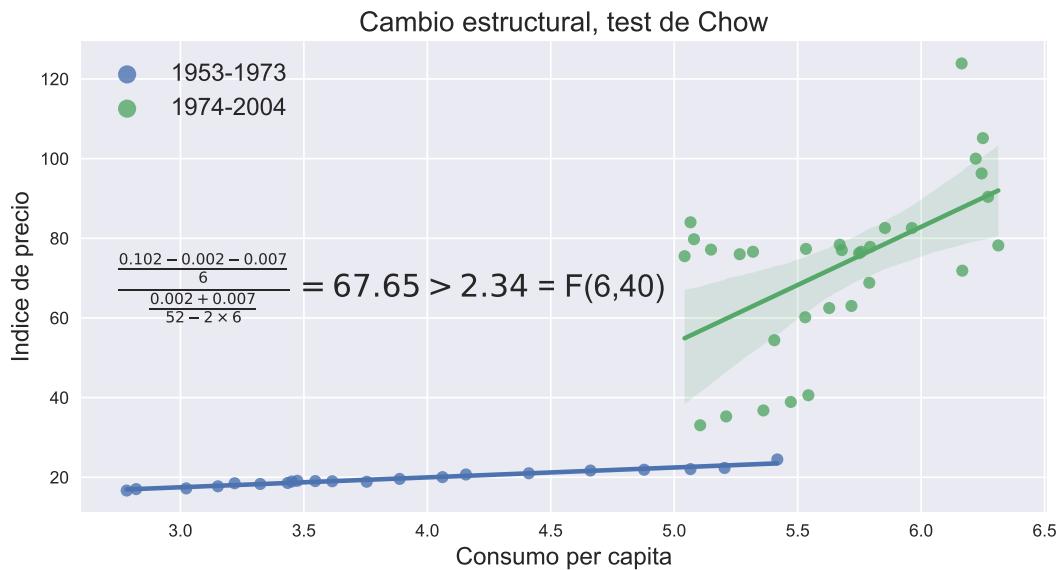


Greene (2012) estima el modelo el consumo per capita de combustible G

$$G_t = \beta_1 + \beta_2 I_t + \beta_3 PG_t + \beta_4 PNC_t + \beta_5 PUC_t + \beta_6 t + \epsilon_t$$

(todas las variables en logaritmo, excepto t) y comprueba si hay un cambio estructural en 1974.

La prueba de Chow confirma que hay un cambio estructural en 1974: los parámetros de 1953-1973 son significativamente distintos a los de 1974-2004.



Limitaciones de la prueba de Chow

La prueba de Chow tiene algunas limitaciones importantes

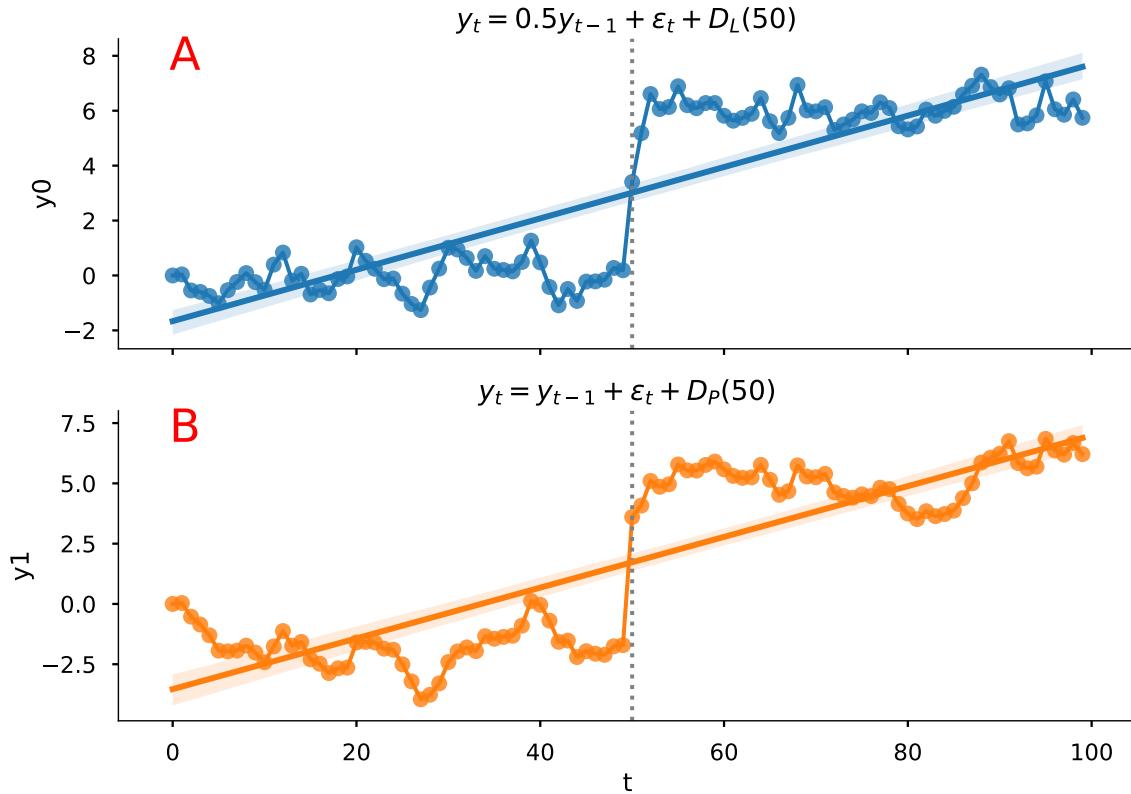
- Asume que el analista sabe la fecha en que ocurre el cambio estructural
- Asume que la varianza de los errores es la misma antes y después del cambio estructural.

6.3 Cambio estructural y raíces unitarias: fecha conocida

Cambio estructural y raíces unitarias

- Cuando se realizan pruebas de raíz unitaria, debe tenerse cuidado si se sospecha que ha ocurrido un cambio estructural.
- Cuando hay cambios estructurales, los estadísticos Dickey-Fuller están sesgados hacia no rechazar la hipótesis de que hay raíz unitaria.

Ejemplos de cambios estructurales



- Los datos de la figura A fueron generados por

$$y_t = 0.5y_{t-1} + \epsilon_t + 3D_L(50)_t \quad (\text{AR}(1) + \text{cambio } \mu)$$

$$D_L(50)_t = I(t \geq 50) \quad (\text{dummy de nivel})$$

- La serie simulada parece tener una tendencia lineal.
- La manera apropiada de estimar este modelo es

$$y_t = \beta_0 + \beta_1 y_{t-1} + \delta D_L(50) + \epsilon_t$$

pero si estimamos

$$y_t = \alpha_0 + \alpha_1 t + \epsilon_t$$

el coeficiente α_1 tendería a ser positivo para capturar el salto en la serie.

- Suponga que por equivocación estimamos

$$y_t - y_{t-1} = c + \gamma y_{t-1} + \epsilon_t$$

lo que sería una caminata aleatoria con deriva si $\gamma = 0$. Entonces:

$$y_t = y_0 + ct + \sum_{i=i}^t \epsilon_i$$

- Esta ecuación describe los datos de manera similar a como lo hace $y_t = \alpha_0 + \alpha_1 t + \epsilon_t$, lo cual nos dice que la regresión Dickey-Fuller sesgaría los resultados hacia $\gamma = 0$, es decir, hacia no rechazar la presencia de una raíz unitaria.

- Perron (1989) demostró este sesgo con simulaciones de Monte Carlo.
- Ahora bien, una serie con raíz unitaria también puede presentar un cambio estructural, como en la figura B, generada por

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t + 5D_P(50)_t && (\text{RW + cambio } \mu) \\ D_P(50)_t &= I(t = 50) && (\text{dummy de impulso}) \end{aligned}$$

- Note que $D_P(t)$ es igual a 1 solo en el período t : cualquier cambio aditivo a una caminata aleatoria tiene un efecto permanente sobre el nivel de la serie.
- La serie simulada parece tener una tendencia lineal.
- De hecho, no es sencillo distinguirla de la figura A a simple vista.
- ¿Cómo saber entonces si una serie con cambio estructural tiene raíz unitaria?
- Una opción, similar a lo que hace la prueba de Chow, es partir la muestra de datos en dos partes, y realizar pruebas de raíz unitaria en cada una por separado
- Lo malo de este procedimiento es que pierde muchos grados de libertad.
- Sería preferible tener un test que utilice toda la muestra

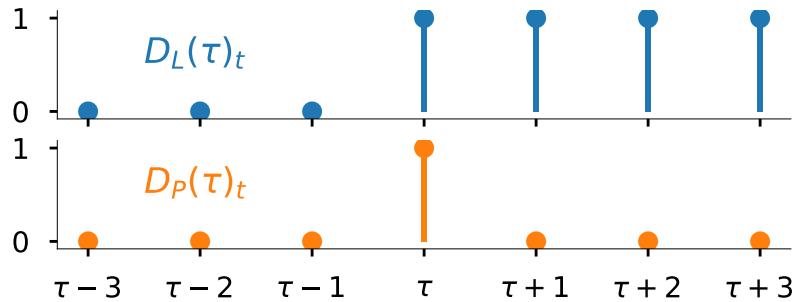
¿Caminata aleatoria o serie estacionaria alrededor de tendencia?

- Perron (1989) desarrolla un procedimiento formal para detectar raíces unitarias en presencia de un cambio estructural ocurrido en $t = \tau$.
- Las hipótesis nula y alternativa son

$$\begin{aligned} y_t &= \alpha_0 + y_{t-1} + \mu_1 D_t^P(\tau) + \epsilon_t && (\text{nula}) \\ y_t &= \alpha_0 + \alpha_2 t + \mu_2 D_t^L(\tau) + \epsilon_t && (\text{alternativa}) \end{aligned}$$

- Es decir, la hipótesis nula es que hay un solo salto en una caminata aleatoria en el período τ , la alternativa es que en esa misma fecha hay un solo salto en el intercepto de una serie estacionaria alrededor de tendencia.

- Observemos que
- $$\sum_{k=1}^t D_p(\tau)_k = D_L(\tau)_t$$



- Integrando la hipótesis nula

$$y_t = y_0 + \alpha_0 t + \mu_1 D_L(\tau)_t + \sum_{k=0}^t \epsilon_k$$

vemos que es equivalente a la alternativa excepto por las perturbaciones.

Las pruebas de cambio estructural de Perron

Modelo A: cambio de intercepto (crash)

$$y_t = \underbrace{\alpha_0 + \mu_1 D_t^P}_{\text{intercepto}} + y_{t-1} + \epsilon_t \quad (\text{nula})$$

$$y_t = \underbrace{\alpha_0 + \mu_2 D_t^L}_{\text{intercepto}} + \alpha_2 t + \epsilon_t \quad (\text{alternativa})$$

Modelo B: cambio de tendencia (changing growth)

$$y_t = \underbrace{\alpha_0 + \mu_2 D_t^L}_{\text{intercepto}} + y_{t-1} + \epsilon_t \quad (\text{nula})$$

$$y_t = \alpha_0 + \underbrace{\alpha_2 t + \mu_3 D_t^T}_{\text{tendencia}} + \epsilon_t \quad (\text{alternativa})$$

donde $D_t^T(\tau) = \max(t - \tau, 0)$

Modelo C: cambio de intercepto y de tendencia

$$y_t = \underbrace{\alpha_0 + \mu_1 D_t^P + \mu_2 D_t^L}_{\text{intercepto}} + y_{t-1} + \epsilon_t \quad (\text{nula})$$

$$y_t = \underbrace{\alpha_0 + \mu_2 D_t^L}_{\text{intercepto}} + \underbrace{\alpha_2 t + \mu_3 D_t^T}_{\text{tendencia}} + \epsilon_t \quad (\text{alternativa})$$

Para implementar la prueba de Perron se siguen 3 pasos:

Paso 1 Se estima la regresión correspondiente al modelo

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + \mu_1 D_t^P + \mu_2 D_t^L + \epsilon_t \quad (\text{A})$$

$$y_t = \alpha_0 + \alpha_2 t + \mu_3 D_t^T + \tilde{y}_t \Rightarrow \tilde{y}_t = \alpha_1 \tilde{y}_{t-1} + \epsilon_t \quad (\text{B})$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + \mu_1 D_t^P + \mu_2 D_t^L + \mu_3 D_t^L + \epsilon_t \quad (\text{C})$$

Paso 2 Si los residuos de la regresión están autocorrelacionados, se estima la regresión ampliada, agregando p rezagos del cambio en la variable:

$$\text{A y C : } \sum_{i=1}^p \beta_i \Delta y_{t-1} \qquad \qquad \text{B : } \sum_{i=1}^p \beta_i \Delta \tilde{y}_{t-1}$$

Paso 3 Se calcula el estadístico t de la hipótesis $\alpha_1 = 1$:

$$t_{\alpha_1} = \frac{\hat{\alpha}_1 - 1}{s.e.(\alpha_1)}$$

y se compara con el valor crítico de Perron.

- Si el estadístico estimado es menor que el valor crítico, se rechaza la hipótesis nula.
- El valor crítico depende de la proporción λ de observaciones que corresponden al periodo anterior al quiebre estructural.

	1953-2004	1953-1973	1974-2004	descripción
intercepto	-12.863	*	-8.349	*
I	1.625	*	0.848	*
PG	-0.054	-0.032	-0.124	*
PNC	-0.083	0.699	*	-0.001
PUC	-0.085	-0.291	*	-0.022
t	-0.014	*	0.010	*
R^2	0.965	0.997	0.953	R^2
S	0.102	0.002	0.007	suma residuos ²

* significativo al 5%

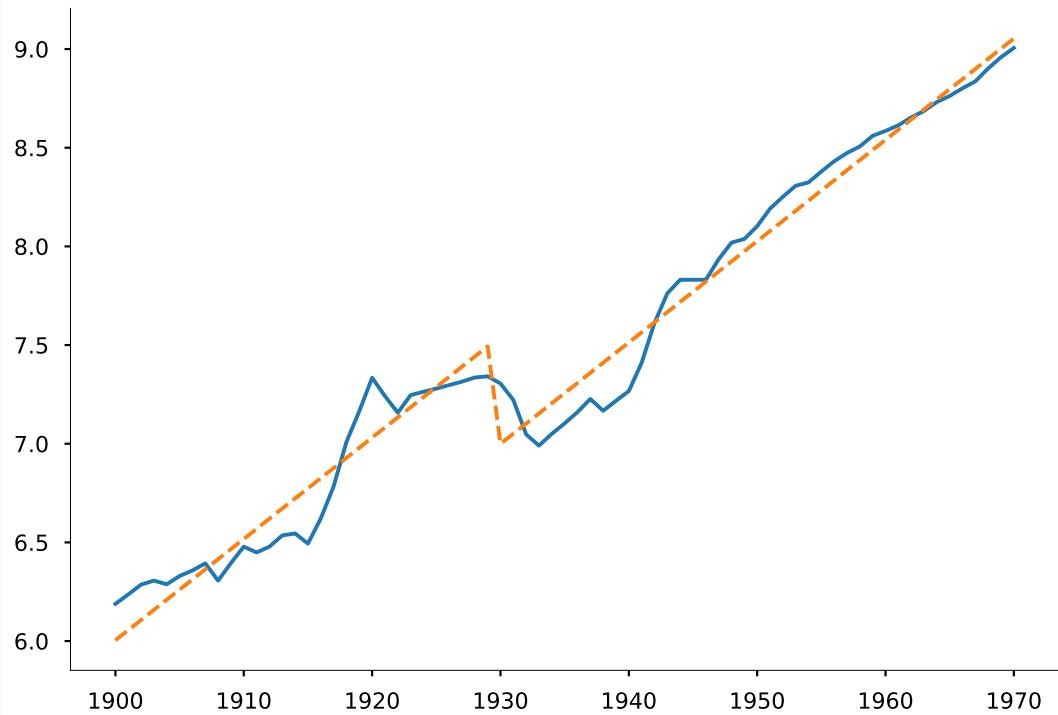
Ejemplo 6-2: Pruebas de cambio estructural

 NelsonPlosserData.dta

 Perron1989.py

Perron analiza los datos de Nelson y Plosser.

Salarios nominales



Al estimar el modelo A encuentran

	<i>t</i>	λ	1%	5%	10%
Real GNP	-5.03	0.34	-4.39	-3.76	-3.46
Nominal GNP	-5.42	0.34	-4.39	-3.76	-3.46
Real per capita GNP	-4.09	0.34	-4.39	-3.76	-3.46
Industrial production	-5.47	0.63	-4.45	-3.76	-3.47
Employment	-4.51	0.49	-4.32	-3.76	-3.46
GNP deflator	-4.04	0.50	-4.32	-3.76	-3.46
Consumer prices	-1.28	0.63	-4.45	-3.76	-3.47
Wages	-5.41	0.42	-4.34	-3.72	-3.44
Money stock	-4.29	0.50	-4.32	-3.76	-3.46

A continuación vemos cómo replicar los resultados del modelo A de Perron, escribiendo un programa de Python.

```
import pandas as pd
import numpy as np
from statsmodels.formula.api import ols

NP = pd.read_stata('NelsonPlosserData.dta')
NP.set_index('year', inplace=True)
NP.index = NP.index.year
```

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + \mu_1 D_t^P + \mu_2 D_t^L + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t$$

```
def Perron_testA(serie, k=8):
    dta = NP[[serie]].dropna()
    dta.rename(columns={serie:'y'}, inplace=True)
    dta['DL'] = (dta.index > 1929)
    dta['DP'] = (dta.index == 1930)
    dta['t'] = np.arange(dta.shape[0])
    dta['Ly'] = dta['y'].shift(1)
    dta['Dy'] = dta['y'].diff(1)
    for j in range(1, k+1):
        dta[f'D{j}y'] = dta['Dy'].shift(j)

    lags = '+'.join(dta.columns[-k:])

    modelo = ols('y ~ DL + DP + t + Ly + ' + lags, dta).fit()
    tval = ((modelo.params - 1)/modelo.bse).round(2)[ 'Ly']
    lda = 1-dta['DL'].mean()
    crit = perron['Modelo A'].loc[np.round(lda, 1)]

    return {'$t$':tval, '$\lambda$':np.round(lda, 2), '1%':crit['1%'], '5%':crit['5%'], '10%':crit['10%']}
```

6.4 Cambio estructural y raíces unitarias: fecha desconocida

¿Cómo saber cuándo se produjo un cambio estructural?

- Un supuesto importante en la prueba de Perron (1989) es que el analista conoce la fecha en que se produjo el (único) cambio estructural.
- No obstante, esto no siempre es factible.
- Zivot y Andrews (1992) proponen una prueba de raíz unitaria similar a la de Perron, pero que asume que el momento del cambio estructural es desconocido.

Las pruebas de cambio estructural de Zivot y Andrews

Una hipótesis nula, tres alternativas

$y_t = \mu + y_{t-1} + \epsilon_t$	(nula)
$y_t = \underbrace{\alpha_0 + \mu_2 D_t^L}_{\text{intercepto}} + \alpha_2 t + \epsilon_t$	(alternativa A)
$y_t = \alpha_0 + \underbrace{\alpha_2 t + \mu_3 D_t^T}_{\text{tendencia}} + \epsilon_t$	(alternativa B)
$y_t = \underbrace{\alpha_0 + \mu_2 D_t^L}_{\text{intercepto}} + \underbrace{\alpha_2 t + \mu_3 D_t^T}_{\text{tendencia}} + \epsilon_t$	(alternativa C)

Para implementar la prueba de Zivot y Andrews se siguen estos pasos:

Paso 1 Se estima la regresión correspondiente al modelo

$$\begin{aligned}
 \textcolor{brown}{A} : y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t & + \mu_2 D_t^L & + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t \\
 \textcolor{brown}{B} : y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t & + \mu_3 D_t^T & + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t \\
 \textcolor{brown}{C} : y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t & + \mu_2 D_t^L + \mu_3 D_t^T & + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t
 \end{aligned}$$

donde los términos D_t^L y D_t^T dependen de la proporción de datos λ anteriores al cambio estructural:

$$D_t^L(\lambda) = I(t > \lambda T) \quad D_t^T = \max(t - \lambda T, 0)$$

Paso 2 Se calcula el estadístico t de la hipótesis $a_1 = 1$:

$$t_{\alpha_1} = \frac{\hat{\alpha}_1 - 1}{s.e.(\hat{\alpha}_1)}$$

Observemos que el valor estimado $\hat{\alpha}$ dependerá de λ ; por ello, escribimos $t_{\alpha_1}(\lambda)$

Paso 3 Se define el punto de quiebre $\hat{\lambda}$ como aquel valor λ que hace más plausible la hipótesis alternativa

$$\hat{\lambda} \equiv \underset{\lambda \in (0,1)}{\operatorname{argmin}} \{t_{\alpha_1}(\lambda)\}$$

Paso 4 Se compara el valor mínimo $t_{\alpha_1}(\hat{\lambda})$ con el valor crítico de Zivot y Andrews. Si el estadístico estimado es menor que el valor crítico, se rechaza la hipótesis nula.

Cuadro 6.1: Valores críticos de Perron

λ	Modelo A			Modelo B			Modelo C		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
0.1	-4.30	-3.68	-3.40	-4.27	-3.65	-3.36	-4.38	-3.75	-3.45
0.2	-4.39	-3.77	-3.47	-4.41	-3.80	-3.49	-4.65	-3.99	-3.66
0.3	-4.39	-3.76	-3.46	-4.51	-3.87	-3.58	-4.78	-4.17	-3.87
0.4	-4.34	-3.72	-3.44	-4.55	-3.94	-3.66	-4.81	-4.22	-3.95
0.5	-4.32	-3.76	-3.46	-4.56	-3.96	-3.68	-4.90	-4.24	-3.96
0.6	-4.45	-3.76	-3.47	-4.57	-3.95	-3.66	-4.88	-4.24	-3.95
0.7	-4.42	-3.80	-3.51	-4.51	-3.85	-3.57	-4.75	-4.18	-3.86
0.8	-4.33	-3.75	-3.46	-4.38	-3.82	-3.50	-4.70	-4.04	-3.69
0.9	-4.27	-3.69	-3.38	-4.26	-3.68	-3.35	-4.41	-3.80	-3.46

Fuente: Perron (1989)

Ejemplo 6-3: Pruebas de cambio estructural (continuación)

 NelsonPlosserData.dta

 Perron1989.py

Zivot y Andrews (1992) también analizan los datos de Nelson y Plosser. Al estimar el modelo A encuentran

	\hat{T}_B	α_1	t
Real GNP	1929	0.267	-5.58
Nominal GNP	1929	0.532	-5.82
Real per capita GNP	1929	0.494	-4.61
Industrial production	1929	0.290	-5.95
Employment	1929	0.651	-4.95
GNP deflator	1929	0.786	-4.12
Consumer prices	1873	0.941	-2.76
Wages	1929	0.660	-5.30
Money stock	1929	0.823	-4.34

Recordemos los valores críticos

1%	5%	10%
-5.34	-4.80	-4.58

A continuación vemos cómo replicar los resultados del modelo A de Zivot y Andrews, escribiendo un programa de Python.

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + \mu_2 D_t^L + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \epsilon_t$$

```
def ZivotAndrewsA(serie, k=8):
    dta = NP[[serie]].dropna()
    dta.rename(columns={serie:'y'}, inplace=True)
    dta['t'] = np.arange(dta.shape[0])
    dta['Ly'] = dta['y'].shift(1)
    dta['Dy'] = dta['y'].diff(1)
```

```

for j in range(1, k+1):
    dta[f'D{j}y'] = dta['Dy'].shift(j)

lags = '+' .join(dta.columns[-k:])

a1vals = pd.Series(0.0, index=dta.index[12:-12])
for tau in a1vals.index:
    dta['DL'] = (dta.index>tau)
    reg = ols('y ~ Ly + t + DL + ' + lags, dta).fit()
    a1vals[tau] = ((reg.params - 1)/reg.bse)[ 'Ly']

tauhat, tval = a1vals.idxmin(), a1vals.min()
dta['DL'] = (dta.index>tauhat)
reg = ols('y ~ Ly + t + DL + ' + lags, dta).fit()

return {r'$\hat{T}_B$':tauhat, r'$\alpha_1$': reg.params[ 'Ly'], r'$t$':tval}

```

Referencias del capítulo 6

- Enders, Walter (2015). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Greene, William H. (2012). *Econometric Analysis*. 7^a ed. Prentice Hall. ISBN: 978-0-13-139538-1.
- Levendis, John D. (2018). *Time Series Econometrics. Learning Through Replication*. Springer. ISBN: 978-3-319-98281-6.
- Perron, Pierre (nov. de 1989). “The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis”. En: *Econometrica* 57.6, pág. 13611401.
- Zivot, Eric y Donald W. K. Andrews (jul. de 1992). “Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis”. En: *Journal of Business & Economic Statistics* 10.3, págs. 251-270.

Capítulo 7

Modelos de estacionalidad

En este capítulo

7.1	Introducción	127
7.2	Modelo ARIMA estacional	129
7.3	Ajuste estacional	134
7.4	Raíces unitarias estacionales	138

7.1 Introducción

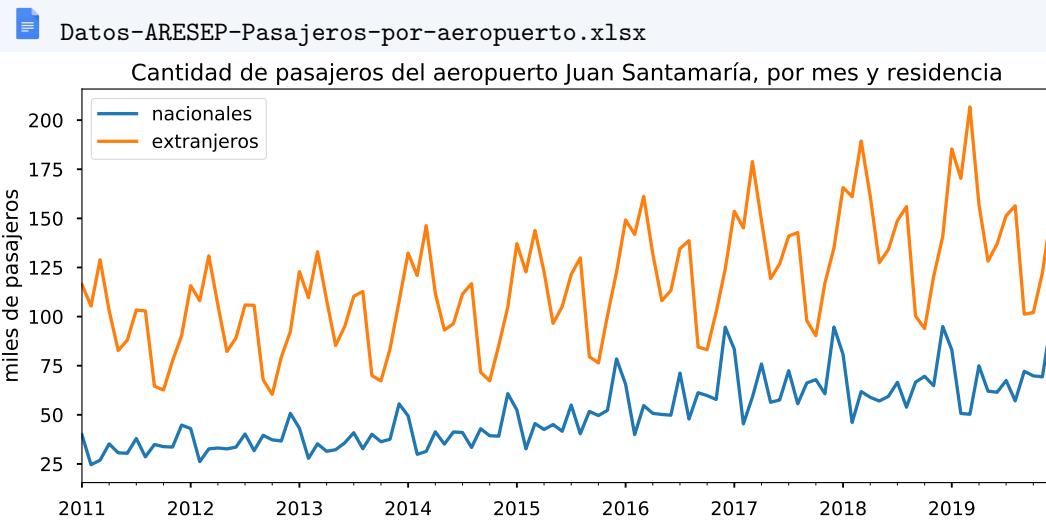
Series con estacionalidad

- En una serie de tiempo, *estacionalidad* es un patrón regular de cambios que se repite cada s períodos.
- Por ejemplo, una serie mensual es estacional si valores altos tienden a ocurrir en meses particulares y los valores bajos tienen a ocurrir en otros meses particulares. En este caso $s = 12$.
- Para series trimestrales estacionales, $s = 4$.
- El primer paso para detectar si una serie es estacional es simplemente graficarla.

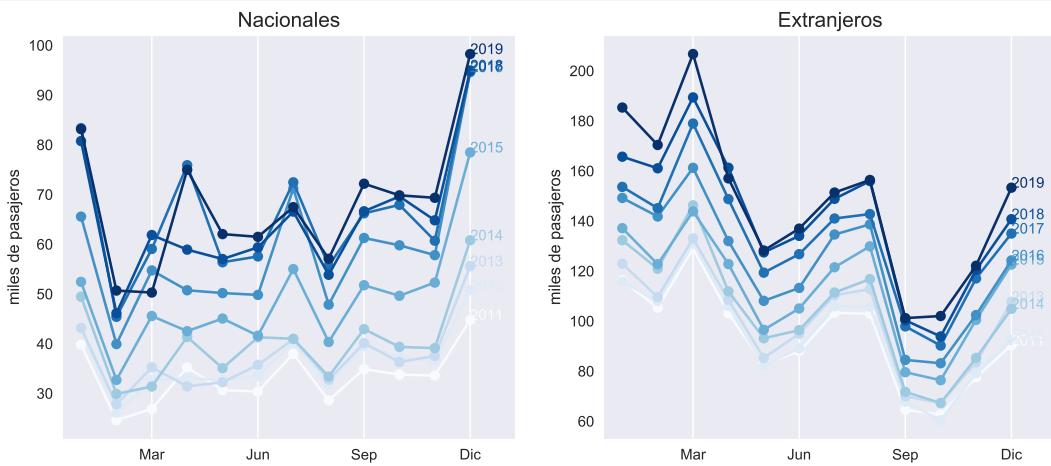


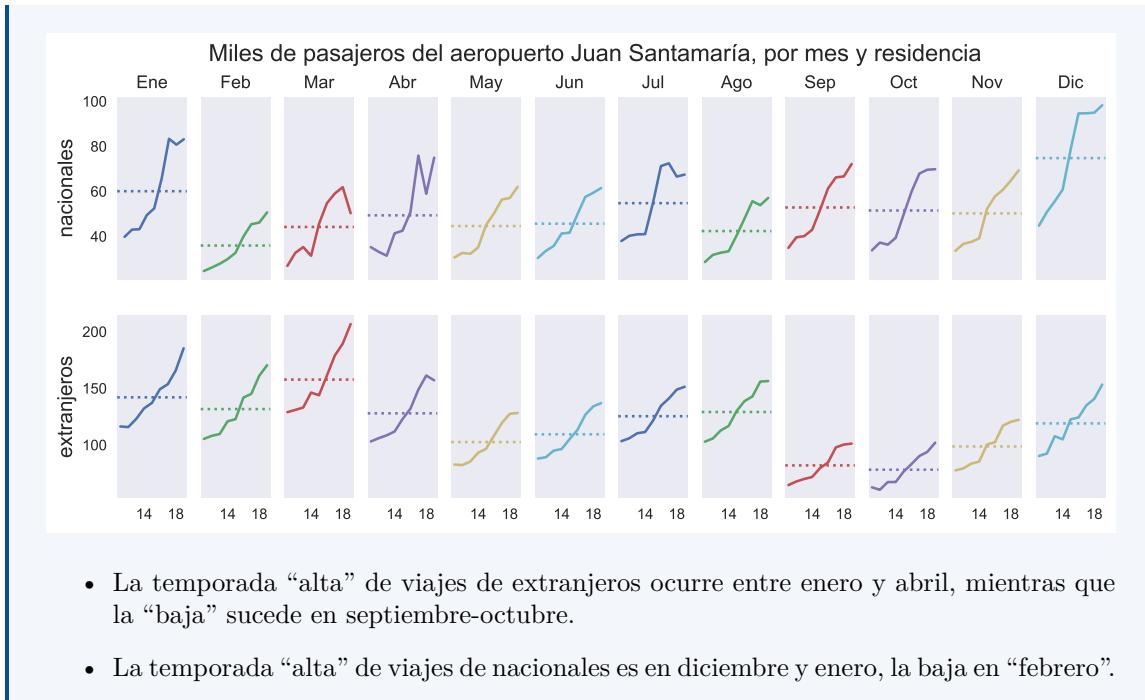
! ¡Estacionalidad no es lo mismo que estacionariedad!

Ejemplo 7-1: Pasajeros del aeropuerto Juan Santamaría



- Un gráfico estacional es similar a un gráfico de tiempo, excepto que los datos se mapean contra la “estación” en la cual se observan.
- Permiten percibir el patrón estacional subyacente más claramente.
- Es especialmente útil para identificar años en que cambian los patrones.





7.2 Modelo ARIMA estacional

Modelo ARIMA estacional

- En un modelo ARIMA estacional, términos AR y MA predicen y_t usando valores de datos y errores con rezagos que son múltiplos de s
- Por ejemplo, con datos trimestrales ($s=4$),
 - un modelo autorregresivo estacional de primer orden usa y_{t-4} para predecir y_t , mientras que uno de segundo orden usa y_{t-4}, y_{t-8} para ello.
 - un modelo de media móvil estacional de primer orden usa ϵ_{t-4} para predecir y_t , mientras que uno de segundo orden utiliza $\epsilon_{t-4}, \epsilon_{t-8}$.
- Por ejemplo, si quisiéramos pronosticar el número de pasajeros extranjeros que viajarán por el SJO en agosto 2020, tiene mucho sentido modelar ese valor en función del número de pasajeros extranjeros que viajaron por SJO en agosto de años anteriores.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
2011	116	105	129	103	83	88	103	103	65	63	78	90
2012	116	108	131	106	82	89	106	106	68	60	79	92
2013	123	110	133	108	85	95	110	113	70	67	84	108
2014	132	121	146	112	93	96	111	117	72	67	85	105
2015	137	123	144	123	97	105	122	130	80	76	101	123
2016	149	142	161	132	108	113	135	139	85	83	102	124
2017	154	145	179	149	119	127	141	143	98	90	117	135
2018	166	161	189	161	127	134	149	156	100	94	120	141
2019	185	170	207	157	128	137	151	156	101	102	122	153

Un modelo estacional puro

	May	Jun	Jul	Ago
2016			y_{t-36-1}	$y_{t-36} = 138.7$
2017			y_{t-24-1}	$y_{t-24} = 142.8$
2018			y_{t-12-1}	$y_{t-12} = 156.0$
2019	$y_{t-3} = 128.2$	$y_{t-2} = 137.0$	$y_{t-1} = 151.4$	$y_t = 156.4$

- En general, un modelo puramente estacional puede representarse por

$$y_t = \varphi_1 y_{t-s} + \varphi_2 y_{t-2s} + \cdots + \varphi_P y_{t-Ps} + \\ + \varepsilon_t + \vartheta_1 \varepsilon_{t-s} + \vartheta_2 \varepsilon_{t-2s} + \cdots + \vartheta_Q \varepsilon_{t-Qs}$$

- o bien, en términos de polinomios de rezagos

$$\underbrace{(1 - \varphi_1 L^s - \cdots - \varphi_P L^{Ps})}_{\Phi(L^s)} y_t = \underbrace{(1 + \vartheta_1 L^s + \cdots + \vartheta_Q L^{Qs})}_{\Theta(L^s)} \varepsilon_t$$

Diferenciación estacional

- La diferenciación estacional se define como la diferencia entre un valor y_t y un valor rezagado un múltiplo de s períodos.

$$\Delta_s y_t = (1 - L^s) y_t = y_t - y_{t-s}$$

- Por ejemplo:

trimestral $\Delta_4 y_t = (1 - L^4) y_t = y_t - y_{t-4}$

mensual $\Delta_{12} y_t = (1 - L^{12}) y_t = y_t - y_{t-12}$

- Nótese que siguiendo esta notación $\Delta y_t = \Delta_1 y_t$. Pero

$$\Delta_2 y_t \neq \Delta^2 y_t$$

⚠ $(1 - L^2) y_t \neq (1 - L)^2 y_t$

$$(1 - L^2) y_t \neq (1 - 2L + L^2) y_t$$

$$y_t - y_{t-2} \neq y_t - 2y_{t-1} + y_{t-2}$$

Un modelo ARIMA para observaciones de la misma estación

- En general, podríamos plantear un modelo ARIMA para observaciones de una sola estación (mes, trimestre)

$$\Phi(L^s) \Delta_s^D y_t = \Theta(L^s) \varepsilon_t$$

- Es usualmente razonable asumir que esta misma relación la cumplen las observaciones de la estación anterior

$$\Phi(L^s) \Delta_s^D y_{t-1} = \Theta(L^s) \varepsilon_{t-1}$$

- En general, los errores de estas relaciones $\varepsilon_t, \varepsilon_{t-1}$ podrían estar correlacionadas, por lo que en principio podemos plantear el modelo ARIMA:

$$\Phi(L) \Delta^d \varepsilon_t = \Theta(L) \epsilon_t$$

donde ϵ_t es ruido blanco.

El modelo SARIMA

- Asumiendo que el proceso ARIMA de cada estación es invertible obtenemos

$$\begin{aligned}\Phi(L^s)\Delta_s^D y_t &= \Theta(L^s)\varepsilon_t \\ \Theta^{-1}(L^s)\Phi(L^s)\Delta_s^D y_t &= \varepsilon_t \\ \Phi(L)\Delta^d\Theta^{-1}(L^s)\Phi(L^s)\Delta_s^D y_t &= \Phi(L)\Delta^d\varepsilon_t \\ &= \Theta(L)\varepsilon_t \\ \Phi(L)\underset{p}{\cancel{\Phi}}(L^s)\underset{P}{\cancel{\Delta}}\underset{d}{\cancel{\Delta}}\underset{D}{\cancel{\Delta}}_s^D y_t &= \Theta(L)\underset{q}{\cancel{\Theta}}(L^s)\underset{Q}{\cancel{\varepsilon}}_t\end{aligned}$$

- Este modelo se denomina $SARIMA(p, d, q) \times (P, D, Q)_s$.

Algunos ejemplos

- $SARIMA(0, 0, 0) \times (0, 1, 0)_4 = SARIMA(0, 1, 0)_4 =$

$$\begin{aligned}\Delta_4 y_t &= \varepsilon_t \\ (1 - L^4)y_t &= \varepsilon_t \\ y_t - y_{t-4} &= \varepsilon_t\end{aligned}$$

- $SARIMA(2, 0, 0)_4 = SAR(2)_4$

$$\begin{aligned}(1 - \varphi_1 L^4 - \varphi_2 L^8)y_t &= \varepsilon_t \\ y_t - \varphi_1 y_{t-4} - \varphi_2 y_{t-8} &= \varepsilon_t\end{aligned}$$

- $SARIMA(0, 0, 3)_4 = SMA(3)_4$

$$\begin{aligned}y_t &= (1 + \vartheta_1 L^4 + \vartheta_2 L^8 + \vartheta_3 L^{12})\varepsilon_t \\ &= \varepsilon_t + \vartheta_1 \varepsilon_{t-4} + \vartheta_2 \varepsilon_{t-8} + \vartheta_3 \varepsilon_{t-12}\end{aligned}$$

- $SARIMA(0, 1, 1) \times (0, 1, 1)_4$

$$\begin{aligned}\Delta\Delta_4 y_t &= (1 + \theta L)(1 + \vartheta L^4)\varepsilon_t \\ (1 - L)(1 - L^4)y_t &= (1 + \theta L)(1 + \vartheta L^4)\varepsilon_t \\ (1 - L - L^4 + L^5)y_t &= (1 + \theta L + \vartheta L^4 + \theta\vartheta L^5)\varepsilon_t \\ y_t - y_{t-1} - y_{t-4} + y_{t-5} &= \varepsilon_t + \theta\varepsilon_{t-1} + \vartheta\varepsilon_{t-4} + \theta\vartheta\varepsilon_{t-5}\end{aligned}$$

- $SARIMA(1, 0, 0) \times (0, 1, 1)_4$

$$\begin{aligned}(1 - \phi L)\Delta_4 y_t &= (1 + \vartheta L^4)\varepsilon_t \\ (1 - \phi L)(1 - L^4)y_t &= (1 + \vartheta L^4)\varepsilon_t \\ (1 - \phi L - L^4 + \phi L^5)y_t &= (1 + \vartheta L^4)\varepsilon_t \\ y_t - \phi y_{t-1} - y_{t-4} + \phi y_{t-5} &= \varepsilon_t + \vartheta\varepsilon_{t-4}\end{aligned}$$

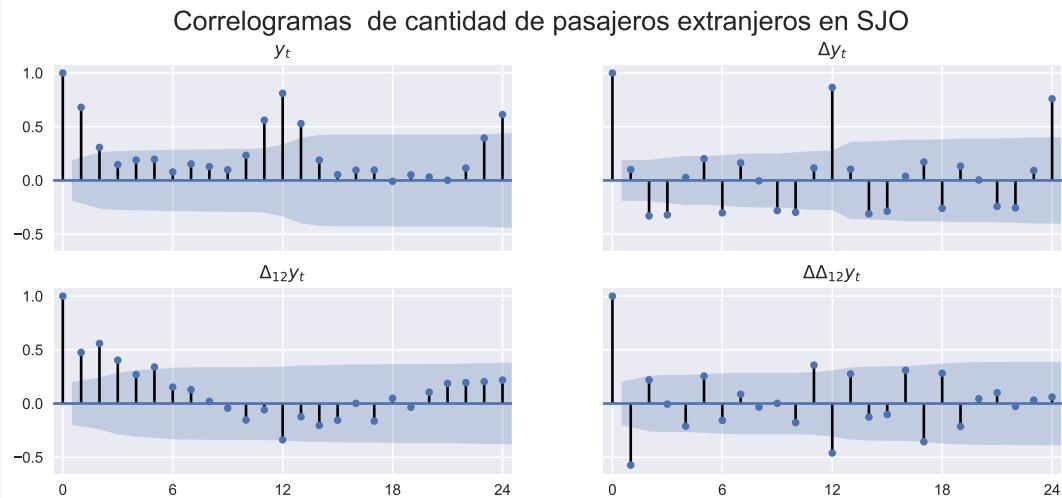
Identificación de un modelo SARIMA

- La identificación de modelos estacionales es más difícil que la identificación de modelos no estacionales por dos razones:

1. Muchas series estacionales exhiben también patrones no estacionales y por lo tanto las FAC y las FACP estimadas contienen ambos patrones.
 2. No hay muchas correlaciones en valores k múltiplos de s . Por ejemplo, en una serie mensual podríamos contar únicamente con $k = 12$, $k = 24$ y $k = 36$
- En la práctica, cuando se tienen dudas, se utilizan herramientas que automatizan esta selección de parámetros a partir de criterios de selección.

Ejemplo 7-2: Estimación de un modelo SARIMA

- El correlograma de la serie de movimientos de pasajeros extranjeros en SJO sugiere que la serie tiene un componente estacional.
- Al usar una herramienta de selección de modelos ([statsmodels.tsa.x13.x13_arima_select_order](#) en Python) se selecciona el modelo SARIMA(0, 1, 1) × (0, 1, 1)₁₂.



Dep. Variable:	0	No. Observations:	108		
Model:	SARIMAX(0, 1, 1)x(0, 1, 1, 12)	Log Likelihood	194.101		
Date:	Sun, 24 May 2020	AIC	-382.202		
Time:	22:03:25	BIC	-374.540		
Sample:	01-31-2011 - 12-31-2019	HQIC	-379.106		
	coef	std err	z	P> z	[0.025 0.975]
ma.L1	-0.6110	0.087	-7.030	0.000	-0.781 -0.441
ma.S.L12	-0.7646	0.117	-6.523	0.000	-0.994 -0.535
sigma2	0.0009	0.000	6.348	0.000	0.001
Ljung-Box (Q):	47.51	Jarque-Bera (JB):	0.23		
Prob(Q):	0.19	Prob(JB):	0.89		
Heteroskedasticity (H):	1.56	Skew:	0.09		
Prob(H) (two-sided):	0.21	Kurtosis:	3.16		

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

- El modelo estimado es

$$(1 - L)(1 - L^{12})y_t = (1 - 0.611L)(1 - 0.765L^{12})\epsilon_t$$

$$y_t - y_{t-1} - y_{t-12} + y_{t-13} = \epsilon_t - 0.611\epsilon_{t-1} - 0.765\epsilon_{t-12} + 0.467\epsilon_{t-13}$$

Los resultados de la tabla anterior muestran que:

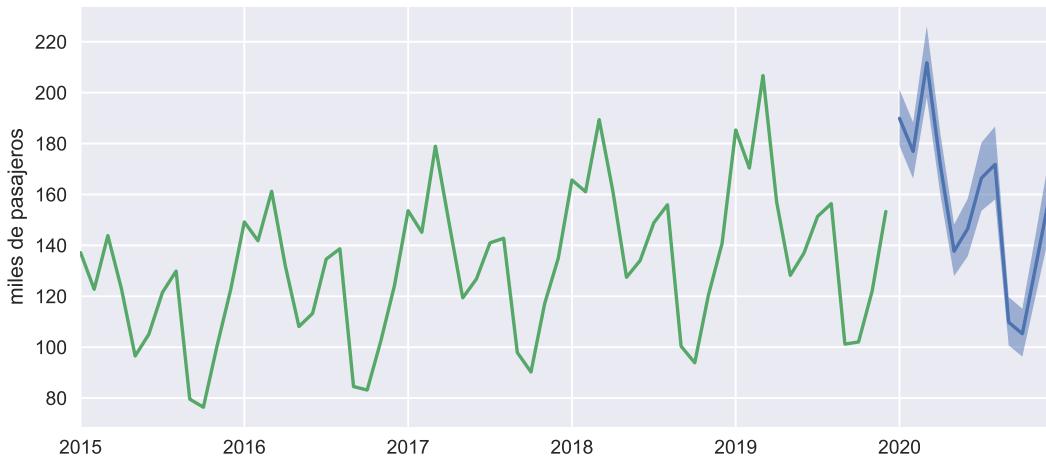
- Los coeficiente estimados son significativos.
- Los residuos del modelo...
 - parecen no estar autocorrelacionados: el valor p del estadístico Q de Lung-Box es 0.19.
 - parecen ser normales: la asimetría es 0.09, la kurtosis es 3.16, y el valor p de la prueba de Jarque-Bera es 0.89.
- Por ello, podemos pensar que el modelo estimado es una buena representación de los datos.

La expresión

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \epsilon_t - 0.611\epsilon_{t-1} - 0.765\epsilon_{t-4} + 0.467\epsilon_{t-5}$$

puede utilizarse recursivamente para generar pronósticos.

Pronóstico de pasajeros extranjeros en aeropuerto SJO



Pronosticando un modelo SARIMA

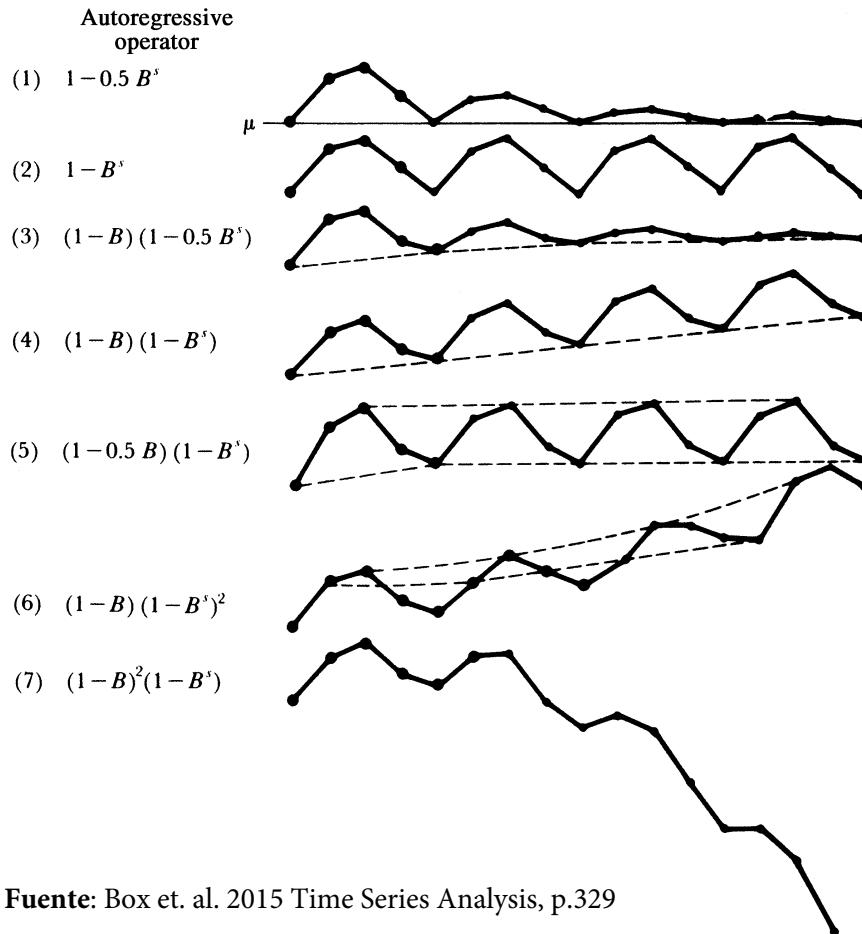
- Si expandimos los polinomios del proceso $\text{SARIMA}(p, d, q) \times (P, D, Q)_s$

$$\underset{p}{\Phi(L)} \underset{P}{\Phi(L^s)} \underset{d}{\Delta^d} \underset{D}{\Delta_s^D} y_t = \underset{q}{\Theta(L)} \underset{Q}{\Theta(L^s)} \epsilon_t$$

el resultado será un polinomio de grado $p + Ps + d + D$ del lado izquierdo y uno de grado $q + Qs$ del lado derecho.

- Para horizontes de más allá de $q + Qs$ períodos, la dinámica de estos pronósticos estará gobernada únicamente por la ecuación en diferencia homogénea

$$\Phi(L)\Phi(L^s)\Delta^d\Delta_s^D \hat{y}_{t+h} = 0$$



Fuente: Box et. al. 2015 Time Series Analysis, p.329

7.3 Ajuste estacional

Ajuste estacional o extracción de señales

- Estos métodos surgieron a mitad del siglo XX para resolver dos problemas:
 1. Remover la tendencia de una serie, con el fin de estudiar apropiadamente su autocorrelación.
 2. Separar las variaciones estacionales de los demás componentes de las series.
- Al inicio, se utilizaban métodos determinísticos, pero luego se fueron abandonando en favor de métodos empíricos basados en promedios móviles

¿Para qué ajustar una serie por estacionalidad?

- El objetivo de los métodos de ajuste estacional es extraer una señal clara de y_t , que permita identificar con facilidad la verdadera evolución de la serie
- Si una serie de tiempo tiene mucho ruido (variabilidad) y un componente estacional fuerte, este ruido afectará a las variaciones inter-periodos o a las variaciones interanuales.
- El ajuste estacional permite hacer comparaciones de un período respecto al anterior, removiendo el ruido que introduciría la presencia de la estacionalidad.

Componentes de las series

Tendencia (T) Es un movimiento de larga duración que se mantiene durante todo el período de observación.

Movimientos cíclicos (C) Son oscilaciones alrededor de la tendencia producidos por períodos alternativos de prosperidad y depresión.

Variación estacional (S) Son los movimientos que se producen dentro del año y que se repiten de un año a otro.

Movimientos irregulares (ϵ) Son las oscilaciones erráticas o accidentales que obedecen a variadas causas. No siguen ningún patrón específico de comportamiento y por tanto son impredecibles.

Descomposición multiplicativa

Multiplicativa

$$y_t = T_t \times C_t \times S_t \times \epsilon_t$$

- Se utiliza cuando se observa que la amplitud del componente estacional varía en forma proporcional al valor de la tendencia.
- En este caso el componente estacional se expresa como un índice y se interpreta como una diferencia relativa. Lo mismo ocurre para el componente irregular.
- La tendencia siempre se mide en las mismas unidades que Y_t

Descomposición aditiva

Aditiva

$$y_t = T_t + C_t + S_t + \epsilon_t$$

- Se utiliza cuando el componente estacional permanece constante, aún si existen cambios en el nivel de tendencia.
- En este caso el componente estacional se expresa en las mismas unidades que Y_t , al igual que ocurre con el componente irregular.
- La tendencia siempre se mide en las mismas unidades que Y_t

Series derivadas

Serie desestacionalizada

Se obtiene cuando a la serie original se le elimina el componente estacional:

$$y_t^{s.a.} = \frac{T_t \times C_t \times S_t \times \epsilon_t}{S_t} = T_t \times C_t \times \epsilon_t$$

Serie tendencia-ciclo

Se obtiene cuando se elimina el componente estacional y el irregular de la serie original:

$$y_t^{t.c.} = \frac{T_t \times C_t \times S_t \times \epsilon_t}{S_t \times \epsilon_t} = T_t \times C_t$$

Métodos actuales para la estimación de componentes

Basados en medias móviles	Basados en modelos
<ul style="list-style-type: none"> • Son empíricos • La estimación de los componentes es local a través de promedio móviles centrados de distintos tamaños • Las revisiones están sujetas a la bondad del pronóstico. 	<ul style="list-style-type: none"> • Proveen medidas de bondad de ajuste sobre la estimación de los componentes. • Permite obtener pronósticos de los componentes. • Minimiza las revisiones. • Minimiza el riesgo de inducción de propiedades espurias.

Línea de tiempo: programas de ajuste estacional

1967 X11, filtros ad-hoc, Julius Shiskin, US Bureau of the Census

1980 X-11 ARIMA, Statistics Canada

1997 X-12 ARIMA, US Bureau of Census

2012 X-13 ARIMA-SEATS, US Bureau of Census [x13as](#)

1992 TRAMO y SEATS (prueba), métodos estocásticos

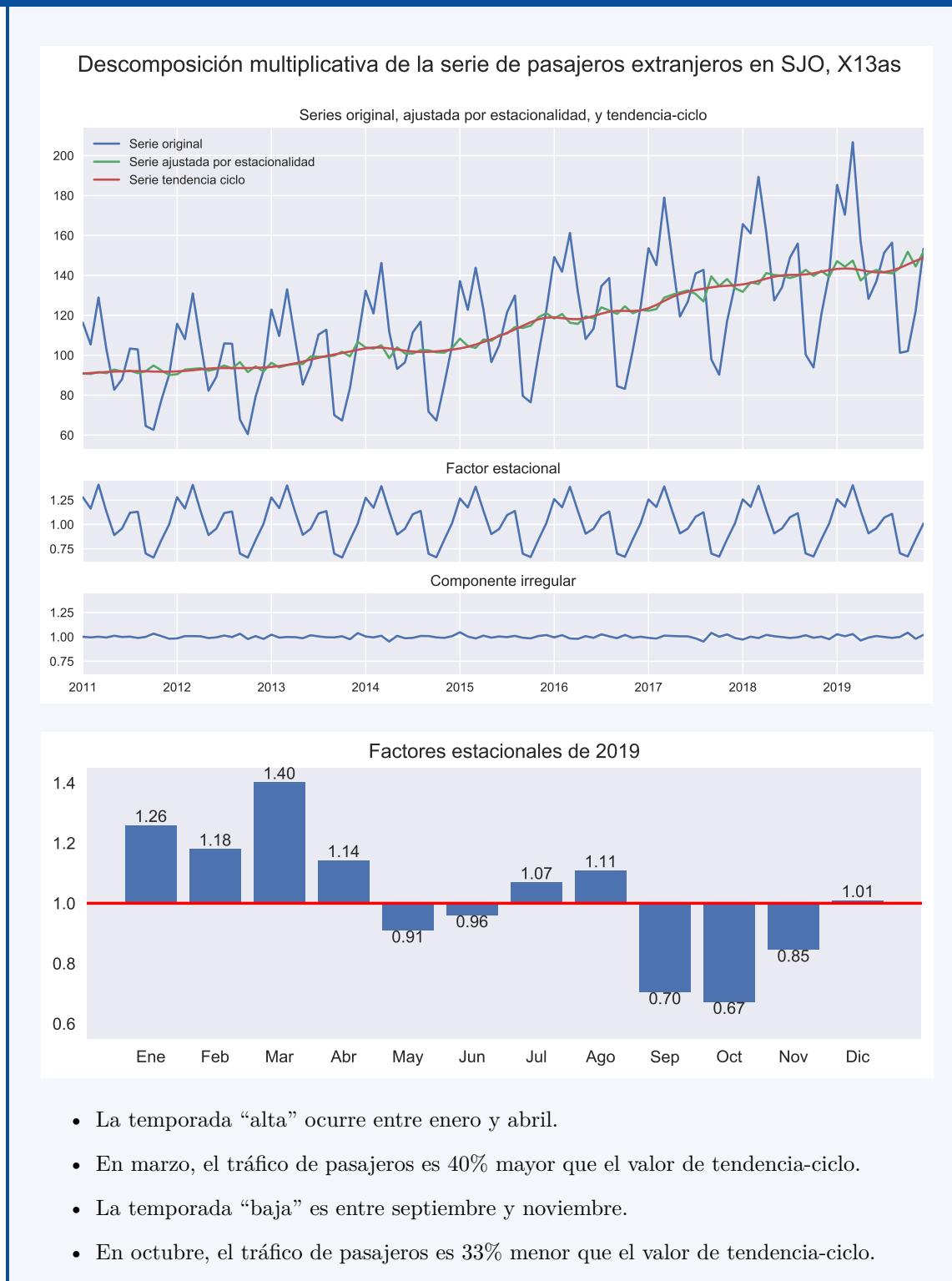
1998 TRAMO y SEATS, Banco de España

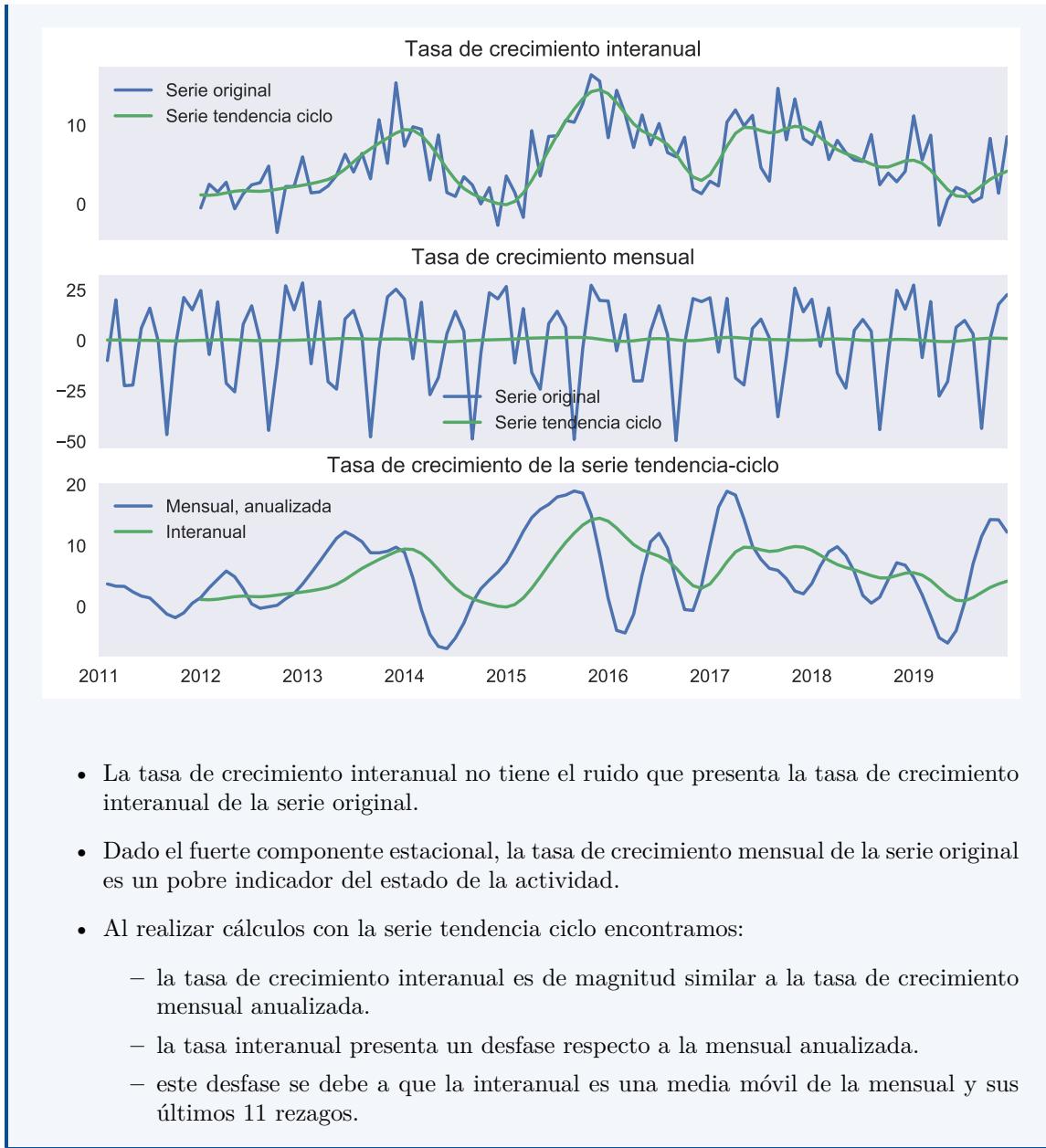
2001 TSW, Banco de España

2012 JDEMETRA+ Eurostat [jdemetra+](#)

2014 TSW+, Banco de España

Ejemplo 7-3: Descomposición de la serie pasajeros extranjeros en SJO, usando X13





- La tasa de crecimiento interanual no tiene el ruido que presenta la tasa de crecimiento interanual de la serie original.
- Dado el fuerte componente estacional, la tasa de crecimiento mensual de la serie original es un pobre indicador del estado de la actividad.
- Al realizar cálculos con la serie tendencia ciclo encontramos:
 - la tasa de crecimiento interanual es de magnitud similar a la tasa de crecimiento mensual anualizada.
 - la tasa interanual presenta un desfase respecto a la mensual anualizada.
 - este desfase se debe a que la interanual es una media móvil de la mensual y sus últimos 11 rezagos.

7.4 Raíces unitarias estacionales

Diferenciación estacional y raíces unitarias

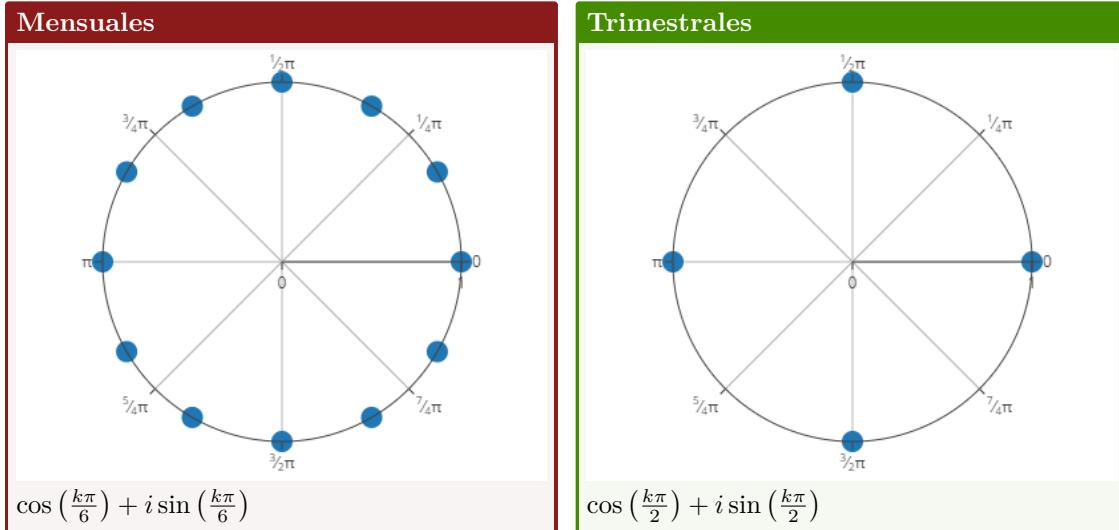
- Si para modelar una serie estacional es necesario aplicar el operador $\Delta_s = (1 - L^s)$, entonces la serie original tiene una raíz unitaria estacional: $1 - 1^s = 0$.
- Pero 1 no es la única raíz de $1 - L^s$: al ser este un polinomio de grado s , tiene exactamente s raíces (no necesariamente todas reales):

$$z_k = e^{\frac{2k\pi}{n}i}, \quad k = 0, 1, \dots, s$$

o bien, usando la fórmula de Euler

$$= \cos\left(\frac{2k\pi}{n}\right) + i \sin\left(\frac{2k\pi}{n}\right)$$

- Todas estas raíces están equidistantes en la circunferencia unitaria.



Pruebas de raíz unitaria estacional

- La literatura provee varias pruebas para detectar raíces unitarias.
- Una de ellas es la prueba de Dickey, Hasza y Fuller (DHF, 1984), la cual es una extensión de la prueba Dickey Fuller:

$$\begin{aligned} y_t &= \rho y_{t-s} + \epsilon_t \\ y_t - y_{t-s} &= (\rho - 1)y_{t-s} + \epsilon_t \\ \Delta_s y_t &= \gamma y_{t-s} + \epsilon_t \end{aligned}$$
- Al igual que en la prueba DF, la hipótesis nula es $\gamma = 0$ contra la alternativa $\gamma < 0$, aunque los valores críticos son distintos.
- Una limitación de esta prueba es que no es apropiada para diagnosticar la presencia de las otras raíces unitarias estacionales.
- Existen otras pruebas de raíces unitarias estacionales
 - La prueba HEGY de Hylleberg, Engle, Granger, y Yoo (1990)
 - La prueba de Kunst (1997)
 - La prueba OCSB, de Osborn, Chui, Smith y Birchenhall (1988)
 - La prueba CH, de Canova y Hansen (1995)
- Sin embargo, no estudiaremos estas pruebas en este curso.

Referencias del capítulo 7

Box, George E. P. y col. (2016). *Time Series Analysis*. 5^a ed. Wiley. ISBN: 978-1-118-67502-1.

Parte II

Métodos para el análisis conjunto de varias series de tiempo

Capítulo 8

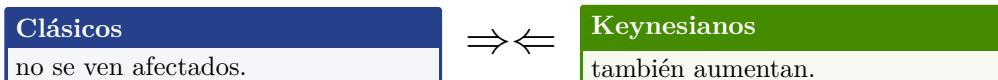
Introducción a la estimación de sistemas de ecuaciones

En este capítulo

8.1 Motivación	143
8.2 El modelo SUR	146

8.1 Motivación

¿Cómo afecta el dinero al producto? Keynesianos vs Clásicos¹
Cuando la oferta de dinero aumenta, el empleo y el producto real...



En palabras de Lucas (1996):

Esta tensión entre dos ideas incompatibles —que cambios en dinero son cambios neutrales de unidades, y que inducen movimientos en empleo y producción de la misma dirección—ha estado en el centro de la teoría monetaria al menos desde Hume (1752).

Un economista al rescate

- Para resolver este problema, un economista estima el modelo

$$y_t = \bar{y} + \alpha_0 m_t + \alpha_1 m_{t-1} + c_0 z_t + c_1 z_{t-1} + u_t$$

donde m es dinero, y es producto, z una variable de control.

- Si $\alpha_0 = \alpha_1 = 0$, los keynesianos estarían en problemas.
- Si $\alpha_0 > 0$, los clásicos estarían en problemas.

¹Basado en Walsh (2010, ch.1)

Una regla de política monetaria

- Suponga que el banco central desea estabilizar el producto alrededor de \bar{y} .
- Para ello fija la oferta de dinero así:

$$\begin{aligned} m_t^* &= \underset{m_t}{\operatorname{argmin}} \mathbb{E} (y_t - \bar{y})^2 \\ &= \underset{m_t}{\operatorname{argmin}} \mathbb{E} (\alpha_0 m_t + \alpha_1 m_{t-1} + c_0 z_t + c_1 z_{t-1} + u_t)^2 \\ &= -\frac{\alpha_1}{\alpha_0} m_{t-1} - \frac{c_1}{\alpha_0} z_{t-1} \end{aligned}$$

donde se supone que el banco central espera $\mathbb{E} z_t = 0$.

- La regla de política sería

$$m_t^* = \pi_1 m_{t-1} + \pi_2 z_{t-1} + \nu_t$$

Otra versión de los hechos

Ahora suponga que el producto real depende solo de cambios *sorpresivos* en la oferta de dinero ν_t :

$$y_t = \bar{y} + d_0 v_t + d_1 z_t + d_2 z_{t-1} + u_t$$

Pero la regla de política implica $v_t = m_t - \pi_1 m_{t-1} - \pi_2 z_{t-1}$. Entonces:

$$\begin{aligned} y_t &= \bar{y} + d_0 [m_t - \pi_1 m_{t-1} - \pi_2 z_{t-1}] + d_1 z_t + d_2 z_{t-1} + u_t \\ &= \bar{y} + d_0 m_t - d_0 \pi_1 m_{t-1} + d_1 z_t + (d_2 - d_0 \pi_2) z_{t-1} + u_t \end{aligned}$$

Un economista en problemas

El economista compara los dos modelos:

<i>keynes</i>	$y_t = \bar{y} + \alpha_0 m_t + \alpha_1 m_{t-1} + c_0 z_t + c_1 z_{t-1} + u_t$
<i>clásico</i>	$y_t = \bar{y} + d_0 m_t - d_0 \pi_1 m_{t-1} + d_1 z_t + (d_2 - d_0 \pi_2) z_{t-1} + u_t$

- La estimación de la regresión no puede distinguir entre las dos hipótesis propuestas: los modelos resultan en regresiones *observacionalmente equivalentes*.
- Los parámetros estimados pueden depender de la regla de política.
- Así, el ejercicio estaría sujeto a la crítica de Lucas (1976): *no podemos predecir qué pasaría si cambia la política, porque el modelo podría no ser invariante a la política misma*.

Un economista con más problemas

- Suponga que el economista se conforma con estimar el modelo

$$y_t = \bar{y} + \alpha_0 m_t + \alpha_1 m_{t-1} + c_0 z_t + c_1 z_{t-1} + u_t$$

y que z_t es el déficit fiscal.

- Si τ es la tasa impositiva media y el gasto público \bar{g} es constante, entonces:

$$z_t = \bar{g} - \tau y_t$$

- En este caso, estimar el modelo por OLS resulta en *estimadores sesgados e inconsistentes!*

Nota: Sesgo de simultaneidad

- Considere el modelo

$$\begin{aligned} C_t &= \alpha + \beta Y_t + \epsilon \\ Y_t &= C_t + I_t \end{aligned}$$

- Esto implica que $Y_t = \frac{\alpha+I_t}{1-\beta} + \frac{\epsilon_t}{1-\beta}$.
- Si se estima la primera ecuación por OLS, la estimación será *inconsistente* porque

$$\text{Cov}(Y_t, \epsilon_t) = \text{Cov}\left(\frac{\epsilon_t}{1-\beta}, \epsilon_t\right) = \frac{1}{1-\beta} \text{Var}(\epsilon_t) \neq 0$$

Un modelo de ecuaciones simultáneas (VAR estructural)

- Dado que en modelos macro las variables son endógenas, es necesario considerar un *sistema de ecuaciones*.

$$\begin{bmatrix} 1 & -\alpha_0 & -c_0 \\ 0 & 1 & 0 \\ \tau & 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ m_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ 0 \\ \bar{g} \end{bmatrix} + \begin{bmatrix} 0 & \alpha_1 & c_1 \\ 0 & \pi_1 & \pi_2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ m_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ \nu_t \\ 0 \end{bmatrix}$$

- Su estimación exige imponer (muchas) restricciones. Por ejemplo, acá imponemos la restricción de que z_t no afecta a m_t en el mismo período.

Resolviendo el sesgo de simultaneidad

Según Sims (1980, pp.14-15)

Debido a que los grandes modelos existentes contienen demasiadas restricciones increíbles, la investigación empírica dedicada a probar teorías macroeconómicas alternativas con demasiada frecuencia procede en un marco de una o pocas ecuaciones. Esta razón es suficiente para que valga la pena investigar la posibilidad de crear grandes modelos en un estilo que no tienda a acumular restricciones tan caprichosamente... *Debe ser factible estimar modelos macro de gran escala como formas reducidas sin restricciones, tratando todas las variables como endógenas.*

Un vector autor-regresivo (VAR)

Así, lo que Sims (1980) propone² es estimar

$$\begin{aligned} y_t &= \bar{y} + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\ m_t &= \bar{m} + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\ z_t &= \bar{z} + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z \end{aligned}$$

- Este es un *modelo reducido*: Las variables y_t, m_t, z_t no interactúan contemporáneamente.
- También es un *modelo SUR*: todas las ecuaciones tienen los mismos regresores; los errores están correlacionados.
- Al ser un modelo SUR con regresores idénticos, *puede estimarse con OLS ecuación por ecuación*.

²El modelo original de Sims es de 6 ecuaciones; acá solo ilustramos la propuesta.

El curso

En la segunda parte de este curso aprenderemos:

- la teoría básica de estimación de sistemas de ecuaciones.
- la distinción entre modelos estructural, recursivo, y reducido.
- la estimación y uso de los modelos VAR y VECM.
- el concepto de integración.
- el uso de herramientas analíticas:
 - la causalidad de Granger,
 - las funciones de impulso respuesta,
 - la descomposición de varianza,
 - los pronósticos.

8.2 El modelo SUR

Introducción

- Hay modelos uniecuacionales que aplican a un grupo de variables relacionadas.
 - Ejemplos:
 - Modelo CAPM
 - Modelos de inversión
- $$r_{it} - r_{ft} = \alpha_i + \beta_i (r_{Mt} - r_{ft}) + \epsilon_{it}$$
- $$I_{it} = \beta_{1i} + \beta_{2i} F_{it} + \beta_{3i} C_{it} + \epsilon_{it}$$
- Como los errores ϵ_{it} de las distintas ecuaciones pueden estar correlacionados, es preferible considerar los modelos de manera conjunta.

Modelo de regresiones aparentemente no relacionadas (SUR)

- En este modelo se presentan un grupo de variables dependientes, pero *NO simultáneas*.
- Cada ecuación puede tener sus propias variables explicativas o éstas pueden ser las mismas para todas las ecuaciones.

El modelo SUR: ecuaciones no simultáneas

Las ecuaciones del sistema son

$$\begin{cases} \mathbf{y}_1 = \mathbf{x}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2 = \mathbf{x}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2 \\ \vdots \\ \mathbf{y}_M = \mathbf{x}_M \boldsymbol{\beta}_M + \boldsymbol{\epsilon}_M \end{cases}$$

que se pueden expresar como

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & 0 & \dots & 0 \\ 0 & \mathbf{x}_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \mathbf{x}_M \end{bmatrix}_{\mathbf{x}} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_M \\ \boldsymbol{\epsilon} \end{bmatrix}$$

El modelo SUR: errores correlacionados

Se asume que perturbaciones de distintas observaciones no están correlacionadas, aunque las perturbaciones de distintas ecuaciones sí pueden estar correlacionados:

$$\mathbb{E} [\epsilon_i \epsilon_j' | \mathbf{x}] = \sigma_{ij} I$$

o bien

$$\mathbb{E}_{\Omega} [\epsilon \epsilon' | \mathbf{x}] = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I & \dots & \sigma_{1M} I \\ \sigma_{21} I & \sigma_{22} I & \dots & \sigma_{2M} I \\ & & \ddots & \\ \sigma_{M1} I & \sigma_{M2} I & \dots & \sigma_{MM} I \end{bmatrix}_{\Sigma \otimes I}$$

Nota: El producto Kronecker

- Si $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ y B son matrices, el producto Kronecker se define por

$$A \otimes B = \begin{bmatrix} aB & bB \\ cB & dB \end{bmatrix}$$

- Algunas propiedades importantes:

$$\begin{aligned} (A \otimes B)' &= A' \otimes B' \\ (A \otimes B)^{-1} &= A^{-1} \otimes B^{-1} \\ (A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \end{aligned}$$

Estimación de un modelo SUR

El modelo SUR

$$\left. \begin{array}{l} \mathbf{Y} = \mathbf{x} \beta + \epsilon \\ \mathbb{E} [\epsilon | \mathbf{x}] = 0 \\ \text{Var} [\epsilon | \mathbf{x}] = \Omega = \Sigma \otimes I \end{array} \right\} \begin{array}{l} \text{supuestos del modelo general} \\ \text{lizado de regresión lineal!} \end{array}$$

puede ser estimado por FGLS:

$$\begin{aligned} \hat{\beta}^{\text{GLS}} &= [\mathbf{x}' \Omega^{-1} \mathbf{x}]^{-1} \mathbf{x}' \Omega^{-1} \mathbf{Y} \\ &= [\mathbf{x}' (\Sigma \otimes I)^{-1} \mathbf{x}]^{-1} \mathbf{x}' (\Sigma \otimes I)^{-1} \mathbf{Y} \end{aligned}$$

Caso especial del modelo SUR

- Si todas las regresiones tienen los mismos regresores, estimar el sistema SUR por GLS es equivalente a estimar ecuación por ecuación con OLS.
- Este resultado justifica que *un VAR sin restricciones se estima ecuación por ecuación con OLS*.

SUR: las regresiones tienen los mismos regresores

En el caso especial $\mathbf{x}_1 = \dots = \mathbf{x}_M = \mathbb{X}$ tenemos

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 & 0 & \dots & 0 \\ 0 & \mathbf{x}_2 & \dots & 0 \\ & \ddots & & \\ 0 & 0 & \dots & \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} \mathbb{X} & 0 & \dots & 0 \\ 0 & \mathbb{X} & \dots & 0 \\ & \ddots & & \\ 0 & 0 & \dots & \mathbb{X} \end{bmatrix} = I \otimes \mathbb{X}$$

y el estimador GLS es

$$\begin{aligned} \hat{\beta}^{\text{GLS}} &= [\mathbf{x}'(\Sigma \otimes I)^{-1} \mathbf{x}]^{-1} \mathbf{x}'(\Sigma \otimes I)^{-1} \mathbf{Y} \\ &= [(I \otimes \mathbb{X})'(\Sigma \otimes I)^{-1}(I \otimes \mathbb{X})]^{-1} (I \otimes \mathbb{X})'(\Sigma \otimes I)^{-1} \mathbf{Y} \\ &= [(I \otimes \mathbb{X}')(\Sigma^{-1} \otimes I)(I \otimes \mathbb{X})]^{-1} (I \otimes \mathbb{X}')(\Sigma^{-1} \otimes I) \mathbf{Y} \\ &= [\Sigma^{-1} \otimes (\mathbb{X}'\mathbb{X})]^{-1} [\Sigma^{-1} \otimes \mathbb{X}'] \mathbf{Y} \\ &= [\Sigma \otimes (\mathbb{X}'\mathbb{X})^{-1}] [\Sigma^{-1} \otimes \mathbb{X}'] \mathbf{Y} \\ &= [I \otimes (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'] \mathbf{Y} \\ &= [I \otimes (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'] \mathbf{Y} \\ &= \begin{bmatrix} (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' & 0 & \dots & 0 \\ 0 & (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} \\ &= \begin{bmatrix} (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' \mathbf{y}_1 \\ (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' \mathbf{y}_2 \\ \vdots \\ (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1^{\text{OLS}} \\ \hat{\beta}_2^{\text{OLS}} \\ \vdots \\ \hat{\beta}_M^{\text{OLS}} \end{bmatrix} \end{aligned}$$

Referencias del capítulo 8

- Greene, William H. (2012). *Econometric Analysis*. 7^a ed. Prentice Hall. ISBN: 978-0-13-139538-1.
- Lucas, Robert E. (1996). “Nobel Lecture: Monetary Neutrality”. En: *Journal of Political Economy* 104.4, págs. 661-682.
- Lucas, Robert Jr (ene. de 1976). “Econometric policy evaluation: A critique”. En: *Carnegie-Rochester Conference Series on Public Policy* 1.1, págs. 19-46.
- Sims, Christopher A. (ene. de 1980). “Macroeconomics and Reality”. En: *Econometrica* 48.1.
- Walsh, Carl E. (2010). *Monetary Theory and Policy*. 3^a ed. MIT Press. ISBN: 978-0262-013772.

Capítulo 9

Sistemas de ecuaciones simultáneas

En este capítulo

9.1 Representación	149
9.2 Identificación	152
9.3 El término de error: plim vs \mathbb{E}	159
9.4 Estimación	160

Introducción

- En Economía, muchas de las teorías se construyen como modelos de ecuaciones simultáneas
- Ejemplos:

Un modelo de oferta y demanda

$$\begin{aligned}q_t^d &= \alpha_0 + \alpha_1 p_t + \alpha_2 x_t + \epsilon_t^d \\q_t^s &= \beta_0 + \beta_1 p_t + \epsilon_t^s \\q_t^d &= q_t^s = q\end{aligned}$$

Un modelo de demanda agregada

$$\begin{aligned}C_t &= \alpha_0 + \alpha_1 Y_t + \epsilon_{1t} \\I_t &= \beta_0 + \beta_1 (Y_t - Y_{t-1}) + \epsilon_{2t} \\Y_t &= C_t + I_t + G_t\end{aligned}$$

Modelo de ecuaciones simultáneas

- En este modelo se presentan un grupo de variables dependientes que se determinan simultáneamente en un sistema de ecuaciones.
- Se asume que existen tantas ecuaciones como variables dependientes en el sistema.

9.1 Representación

Modelo en forma estructural

- Hay K variables exógenas, M endógenas, y M ecuaciones.
- Se cuenta con T observaciones de cada variable.

$$\left\{ \begin{array}{l} \gamma_{11}y_{t1} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} = \epsilon_{t1} \\ \gamma_{12}y_{t1} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} = \epsilon_{t2} \\ \vdots \\ \gamma_{1M}y_{t1} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} = \epsilon_{tM} \end{array} \right.$$

Ecuaciones en forma estructural: Matrices

Escribimos las ecuaciones como $y_t' \Gamma + x_t' B = \epsilon_t'$:

$$\begin{bmatrix} y_1 & \dots & y_M \\ & \ddots & \\ y_t' & & \\ \gamma_{M1} & \dots & \gamma_{MM} \\ \Gamma \end{bmatrix}_t \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1M} \\ & \ddots & \\ \gamma_{M1} & \dots & \gamma_{MM} \end{bmatrix} + \begin{bmatrix} x_1 & \dots & x_K \\ & \ddots & \\ x_t' & & \\ \beta_{K1} & \dots & \beta_{KM} \\ B \end{bmatrix}_t \begin{bmatrix} \beta_{11} & \dots & \beta_{1M} \\ & \ddots & \\ \beta_{K1} & \dots & \beta_{KM} \end{bmatrix} = \begin{bmatrix} \epsilon_1 & \dots & \epsilon_M \\ & \ddots & \\ \epsilon_t' & & \end{bmatrix}_t$$

En las matrices Γ y B , cada columna corresponde a una ecuación, y cada fila a una variable.

Modelo en forma estructural: Matrices

Juntando todas las observaciones $y_t' \Gamma + x_t' B = \epsilon_t'$ obtenemos

$$\begin{array}{l} y_1' \Gamma + x_1' B = \epsilon_1' \\ y_2' \Gamma + x_2' B = \epsilon_2' \\ \vdots \\ y_T' \Gamma + x_T' B = \epsilon_T' \end{array} \Rightarrow \begin{bmatrix} y_1' \Gamma \\ y_2' \Gamma \\ \vdots \\ y_T' \Gamma \end{bmatrix} + \begin{bmatrix} x_1' B \\ x_2' B \\ \vdots \\ x_T' B \end{bmatrix} = \begin{bmatrix} \epsilon_1' \\ \epsilon_2' \\ \vdots \\ \epsilon_T' \end{bmatrix} \Rightarrow \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix} \Gamma + \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix} B = \begin{bmatrix} \epsilon_1' \\ \epsilon_2' \\ \vdots \\ \epsilon_T' \end{bmatrix}$$

Modelo en forma reducida

Postmultiplicando la forma estructural por Γ^{-1} tenemos

$$\begin{array}{ll} y_t' \Gamma + x_t' B = \epsilon_t' & Y \Gamma + X B = E \\ y_t' \Gamma \Gamma^{-1} + x_t' B \Gamma^{-1} = \epsilon_t' \Gamma^{-1} & Y \Gamma \Gamma^{-1} + X B \Gamma^{-1} = E \Gamma^{-1} \\ y_t' - x_t' \Pi = \nu_t' & Y - X \Pi = V \\ y_t' = x_t' \Pi + \nu_t' & Y = X \Pi + V \end{array}$$

donde hemos definido $\Pi \equiv -B\Gamma^{-1}$ como los parámetros reducidos del sistema, y $v_t' \equiv \epsilon_t' \Gamma^{-1}$.

Forma Reducida

$$Y = X \Pi + V$$

Ejemplo 9-1: Forma estructural y reducida del modelo de oferta y demanda**Un modelo de oferta y demanda**

$$\begin{cases} q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 x_t + \epsilon_t^d \\ q_t^s = \beta_0 + \beta_1 p_t + \epsilon_t^s \\ q_t^d = q_t^s = q \end{cases}$$

Su forma estructural es

$$\begin{bmatrix} q_t & p_t \\ y'_t & \Gamma \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \end{bmatrix} + \begin{bmatrix} 1 & x_t \\ x'_t & B \end{bmatrix} \begin{bmatrix} -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \end{bmatrix} = \begin{bmatrix} \epsilon_t^d & \epsilon_t^s \\ \epsilon'_t & \endbmatrix$$

Note que *cada columna* de Γ y de B corresponden a una ecuación del modelo.

$$\text{Así, } \Gamma = \begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \end{bmatrix} \Rightarrow \Gamma^{-1} = \frac{1}{\beta_1 - \alpha_1} \begin{bmatrix} \beta_1 & 1 \\ -\alpha_1 & -1 \end{bmatrix}$$

Los parámetros reducidos son:

$$\Pi = -B\Gamma^{-1} = \frac{1}{\beta_1 - \alpha_1} \begin{bmatrix} \alpha_0 & \beta_0 \\ \alpha_2 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 & 1 \\ -\alpha_1 & -1 \end{bmatrix} = \frac{1}{\beta_1 - \alpha_1} \begin{bmatrix} \alpha_0\beta_1 - \alpha_1\beta_0 & \alpha_0 - \beta_0 \\ \alpha_2\beta_1 & \alpha_2 \end{bmatrix}$$

y los shocks reducidos son:

$$\nu'_t = \epsilon'_t \Gamma^{-1} = \frac{1}{\beta_1 - \alpha_1} [\epsilon_t^d \quad \epsilon_t^s] \begin{bmatrix} \beta_1 & 1 \\ -\alpha_1 & -1 \end{bmatrix} = \frac{1}{\beta_1 - \alpha_1} [\beta_1 \epsilon_t^d - \alpha_1 \epsilon_t^s \quad \epsilon_t^d - \epsilon_t^s]$$

por lo que la forma reducida es:

$$\begin{bmatrix} q_t & p_t \\ y'_t & \Pi \end{bmatrix} = \begin{bmatrix} 1 & x_t \\ x'_t & \endbmatrix} \begin{bmatrix} \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{\beta_1 - \alpha_1} & \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} \\ \frac{\alpha_2\beta_1}{\beta_1 - \alpha_1} & \frac{\alpha_2}{\beta_1 - \alpha_1} \end{bmatrix} + \begin{bmatrix} \frac{\beta_1 \epsilon_t^d - \alpha_1 \epsilon_t^s}{\beta_1 - \alpha_1} & \frac{\epsilon_t^d - \epsilon_t^s}{\beta_1 - \alpha_1} \\ \nu'_t & \end{bmatrix}$$

La forma reducida corresponde a la cantidad y precio de equilibrio:

$$\begin{cases} q_t^* = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2\beta_1}{\beta_1 - \alpha_1} x_t + \frac{\beta_1 \epsilon_t^d - \alpha_1 \epsilon_t^s}{\beta_1 - \alpha_1} \\ p_t^* = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} x_t + \frac{\epsilon_t^d - \epsilon_t^s}{\beta_1 - \alpha_1} \end{cases}$$

A partir de la forma reducida, es fácil calcular el efecto de shocks o de cambios en variables exógenas sobre las endógenas. Por ejemplo:

$$\frac{\partial q_t^*}{\partial x_t} = \pi_{21} = \frac{\alpha_2\beta_1}{\beta_1 - \alpha_1} \quad \frac{\partial p_t^*}{\partial \epsilon_t^s} = \frac{-1}{\beta_1 - \alpha_1}$$

Nótese que a partir de la forma reducida, el efecto de una variable exógena es observable (es un parámetro reducido), no así el efecto de un shock estructural (está combinado con los demás shocks estructurales en el shock reducido).

Ejemplo 9-2: Forma estructural y reducida del modelo de demanda agregada

Un modelo de demanda agregada

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \epsilon_{1t} \\ I_t = \beta_0 + \beta_1 (Y_t - Y_{t-1}) + \epsilon_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

Su forma estructural es

$$\begin{bmatrix} 1 & 0 & -\alpha_1 \\ 0 & 1 & -\beta_1 \\ -1 & -1 & 1 \end{bmatrix}_{\Gamma'} \begin{bmatrix} C_t \\ I_t \\ Y_t \end{bmatrix}_{y_t} + \begin{bmatrix} -\alpha_0 & 0 & 0 \\ -\beta_0 & \beta_1 & 0 \\ 0 & 0 & -1 \end{bmatrix}_{B'} \begin{bmatrix} 1 \\ Y_{t-1} \\ G_t \end{bmatrix}_{x_t} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \end{bmatrix}_{\epsilon_t}$$

Como tomamos la transpuesta, note que *cada fila* de Γ' y de B' corresponden a una ecuación del modelo.

Así, $\Gamma' = \begin{bmatrix} 1 & 0 & -\alpha_1 \\ 0 & 1 & -\beta_1 \\ -1 & -1 & 1 \end{bmatrix} \Rightarrow \Gamma'^{-1} = \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} 1-\beta_1 & \alpha_1 & \beta_1 \\ \beta_1 & 1-\alpha_1 & \beta_1 \\ 1 & 1 & 1 \end{bmatrix}$

Los parámetros reducidos son:

$$\begin{aligned} \Pi' &= -\Gamma'^{-1} B' = \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} 1-\beta_1 & \alpha_1 & \beta_1 \\ \beta_1 & 1-\alpha_1 & \beta_1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 & 0 & 0 \\ \beta_0 & -\beta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} \alpha_0 - \alpha_0\beta_1 + \alpha_1\beta_0 & -\alpha_1\beta_1 & \alpha_1 \\ \alpha_0\beta_1 + \beta_0 - \alpha_1\beta_0 & \alpha_1\beta_1 - \beta_1 & \beta_1 \\ \alpha_0 + \beta_0 & -\beta_1 & 1 \end{bmatrix} \end{aligned}$$

y los shocks reducidos son:

$$\begin{aligned} \nu_t &= \Gamma'^{-1} \epsilon_t = \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} 1-\beta_1 & \alpha_1 & \beta_1 \\ \beta_1 & 1-\alpha_1 & \beta_1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \end{bmatrix} \\ &= \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} (1-\beta_1)\epsilon_{1t} + \alpha_1\epsilon_{2t} \\ \beta_1\epsilon_{1t} + (1-\alpha_1)\epsilon_{2t} \\ \epsilon_{1t} + \epsilon_{2t} \end{bmatrix} \end{aligned}$$

por lo que la forma reducida es:

$$\begin{bmatrix} C_t \\ I_t \\ Y_t \end{bmatrix}_{y_t} = \frac{1}{1-\alpha_1-\beta_1} \begin{bmatrix} \alpha_0 - \alpha_0\beta_1 + \alpha_1\beta_0 & -\alpha_1\beta_1 & \alpha_1 \\ \alpha_0\beta_1 + \beta_0 - \alpha_1\beta_0 & \alpha_1\beta_1 - \beta_1 & \beta_1 \\ \alpha_0 + \beta_0 & -\beta_1 & 1 \end{bmatrix}_{\Pi'} \begin{bmatrix} 1 \\ Y_{t-1} \\ G_t \end{bmatrix}_{x_t} + \frac{(1-\beta_1)\epsilon_{1t} + \alpha_1\epsilon_{2t}}{1-\alpha_1-\beta_1} \begin{bmatrix} \beta_1\epsilon_{1t} + (1-\alpha_1)\epsilon_{2t} \\ \epsilon_{1t} + \epsilon_{2t} \\ \nu_t \end{bmatrix}$$

La forma reducida corresponde al consumo, inversión, e ingreso de equilibrio:

$$\begin{cases} C_t^* = \frac{\alpha_0 - \alpha_0\beta_1 + \alpha_1\beta_0}{1-\alpha_1-\beta_1} - \frac{\alpha_1\beta_1}{1-\alpha_1-\beta_1} Y_{t-1} + \frac{\alpha_1}{1-\alpha_1-\beta_1} G_t + \frac{(1-\beta_1)\epsilon_{1t} + \alpha_1\epsilon_{2t}}{1-\alpha_1-\beta_1} \\ I_t^* = \frac{\alpha_0\beta_1 + \beta_0 - \alpha_1\beta_0}{1-\alpha_1-\beta_1} + \frac{\alpha_1\beta_1 - \beta_1}{1-\alpha_1-\beta_1} Y_{t-1} + \frac{\beta_1}{1-\alpha_1-\beta_1} G_t + \frac{\beta_1\epsilon_{1t} + (1-\alpha_1)\epsilon_{2t}}{1-\alpha_1-\beta_1} \\ Y_t^* = \frac{\alpha_0 + \beta_0}{1-\alpha_1-\beta_1} - \frac{\beta_1}{1-\alpha_1-\beta_1} Y_{t-1} + \frac{1}{1-\alpha_1-\beta_1} G_t + \frac{\epsilon_{1t} + \epsilon_{2t}}{1-\alpha_1-\beta_1} \end{cases}$$

Así, el multiplicador del gasto público es:

$$\frac{\partial Y_t^*}{\partial G_t} = \pi_{33} = \frac{1}{1-\alpha_1-\beta_1}$$

De nuevo, observe que los shocks reducidos son combinaciones lineales de los shocks estructurales.

9.2 Identificación

Identificación

- El problema de *identificación* en ecuaciones simultáneas se refiere a cómo obtener los parámetros *estructurales* B, Γ, Σ a partir de los parámetros *reducidos* Π, Ω .

$$\left. \begin{array}{l} \Pi = -B\Gamma^{-1} \\ \Omega = (\Gamma^{-1})' \Sigma \Gamma^{-1} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} B=? \\ \Gamma=? \\ \Sigma=? \end{array} \right.$$

- No es un problema de estimación*, sino de resolución de un sistema de ecuaciones no lineales.

Identificación**(cont'n)**

- Considere la matriz $F \neq I$, y defina $\tilde{\Gamma} = \Gamma F$ y $\tilde{B} = BF$. Entonces

$$\begin{array}{ll} Y\Gamma + XB = E & \text{(estructura verdadera)} \\ Y\tilde{\Gamma} + X\tilde{B} = \tilde{E} & \text{(estructura falsa)} \end{array}$$

$$\begin{aligned} \hat{\Pi} &= -\tilde{B}\tilde{\Gamma}^{-1} \\ &= -BFF^{-1}\Gamma^{-1} \\ &= -B\Gamma^{-1} \\ &= \Pi \end{aligned}$$

- Decimos que las estructuras son *observacionalmente equivalentes*.

Identificación: contando parámetros

	parámetros estructurales	parámetros reducidos
Γ	M^2	
B	KM	Π
Σ	$\frac{M(M+1)}{2}$	Ω
Total:	$\frac{M}{2}(1 + 2K + 3M)$	Total: $\frac{M}{2}(1 + 2K + M)$

Para identificar los parámetros B, Γ, Σ a partir de Π, Ω tenemos:

$$\frac{\frac{M}{2}(1 + 2K + 3M)}{\# \text{ de incógnitas}} - \frac{\frac{M}{2}(1 + 2K + M)}{\# \text{ de ecuaciones}} = \frac{M^2}{\text{exceso de parámetros}}$$

Identificación de la ecuación j

Dado que tenemos más parámetros estructurales que reducidos, es necesario tener información no

- | | | |
|-----------|---|---|
| muestral: | <ul style="list-style-type: none"> • normalizaciones • identidades • exclusiones | <ul style="list-style-type: none"> • restricciones lineales • restricciones en varianza |
|-----------|---|---|

Identificación via restricciones lineales por ecuación

La forma estructural puede escribirse como $Az_t = \epsilon_t$:

$$\begin{bmatrix} \gamma_{11} & \dots & \gamma_{M1} & \beta_{11} & \dots & \beta_{K1} \\ \gamma_{12} & \dots & \gamma_{M2} & \beta_{12} & \dots & \beta_{K2} \\ & & \vdots & & & \\ \gamma_{1M} & \dots & \gamma_{MM} & \beta_{1M} & \dots & \beta_{KM} \end{bmatrix} \begin{bmatrix} y_{t1} \\ \vdots \\ y_{tM} \\ x_{t1} \\ \vdots \\ x_{tK} \end{bmatrix} = \begin{bmatrix} \epsilon_{t1} \\ \epsilon_{t2} \\ \vdots \\ \epsilon_{tM} \end{bmatrix}$$

Algunos de los parámetros de la fila j : $[\gamma_{1j} \dots \gamma_{Mj} \beta_{1j} \dots \beta_{Kj}]$ están restringidos porque:

- una variable endógena está normalizada, ej: $\gamma_{jj} = 1$
- alguna variable está excluida, ej: $\gamma_{3j} = 0$ o bien $\beta_{2j} = 0$
- dos variables tienen el mismo coeficiente, ej: $\beta_{2j} = \beta_{3j}$

Restricción de normalización

La más sencilla de las restricciones es la *normalización*: en cada ecuación, el parámetro de una variable (usualmente endógena) es uno.

Ejemplos:

Un modelo de oferta y demanda	Un modelo de demanda agregada
$1q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 x_t + \epsilon_t^d$ $1q_t^s = \beta_0 + \beta_1 p_t + \epsilon_t^s$ $q_t^d = q_t^s$	$1C_t = \alpha_0 + \alpha_1 Y_t + \epsilon_{1t}$ $1I_t = \beta_0 + \beta_1 (Y_t - Y_{t-1}) + \epsilon_{2t}$ $Y_t = C_t + I_t + G_t$

Restricción de exclusión

La *exclusión* se refiere a que alguna variable del modelo no aparece en cierta ecuación (es decir, su coeficiente es cero en esa ecuación).

Ejemplos:

Un modelo de oferta y demanda	Un modelo de demanda agregada
$q_t^d = \alpha_0 + \alpha_1 p_t + \alpha_2 x_t + \epsilon_t^d$ $q_t^s = \beta_0 + \beta_1 p_t + 0x_t + \epsilon_t^s$ $q_t^d = q_t^s$	$C_t = \alpha_0 + \alpha_1 Y_t + 0Y_{t-1} + 0G_t + 0I_t + \epsilon_{1t}$ $I_t = \beta_0 + \beta_1 (Y_t - Y_{t-1}) + 0G_t + 0C_t + \epsilon_{2t}$ $Y_t = C_t + I_t + G_t$

Restricción de combinación lineal de parámetros

En este caso, una *combinación lineal de parámetros* es conocida. El caso más sencillo es cuando dos parámetros son iguales.

Ejemplo:

Un modelo de demanda agregada
$C_t = \alpha_0 + \alpha_1 Y_t + \epsilon_{1t}$ $I_t = \beta_0 + \beta_1 Y_t + (-\beta_1)Y_{t-1} + \epsilon_{2t}$ $Y_t = C_t + I_t + G_t$

es decir, en la ecuación de inversión la suma de los parámetros de Y_t y de Y_{t-1} debe ser igual a cero.

Nota: El rango de una matriz

- El *rango de una matriz A* de tamaño $M \times N$ se denota por $\text{rango}[A]$ y se define como el número de filas (o columnas) que son linealmente independientes.
- Necesariamente, se cumple que

$$\text{rango}[A] \leq \min \{M, N\}$$

- Si $\text{rango}[A] = M$, decimos que A tiene *rango fila* completo.

- Si $\text{rango}[A] = N$, decimos que A tiene *rango columna* completo.

La condición de rango

Sea \tilde{A}_j la matriz formada por aquellas columnas de A en las que la ecuación j tiene restricciones.

Condición de rango

La ecuación j está identificada si y solo si la matriz \tilde{A}_j tiene rango fila completo; es decir

$$\text{rango} [\tilde{A}_j] = M$$

Esta condición es necesaria y suficiente.

La condición de orden

Note que para que se cumpla la condición de rango, es necesario que \tilde{A}_j tenga al menos M columnas.

Condición de orden

Para que la ecuación j esté identificada, es necesario que el número de restricciones en tal ecuación sea mayor o igual al número de variables endógenas del sistema (= al número de ecuaciones).

Esta condición es necesaria *pero no suficiente*.

Clasificación de ecuaciones

	Condición de orden	Condición de rango
Sobre-identificada	Se cumple con desigualdad	Se cumple
Exactamente identificada	Se cumple con igualdad	Se cumple
Sub-identificada		alguna condición no se cumple

- Se dice que está *identificada* sólo si está “sobre-identificada” o “exactamente-identificada”
- *Solo las ecuaciones identificadas pueden ser estimadas*

Ejemplo 9-3: Identificación en el modelo de oferta y la demanda

En el ejemplo 9.1 encontramos esta forma estructural para el modelo de oferta y demanda

$$[q_t \ p_t] \begin{bmatrix} 1 & -1 \\ -\alpha_1 & -\beta_1 \end{bmatrix} + [1 \ x_t] \begin{bmatrix} -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \end{bmatrix} = [\epsilon_t^d \ \epsilon_t^s]$$

que puede escribirse también como

$$Az_t = \left[\begin{array}{cc|cc} 1 & -\alpha_1 & -\alpha_0 & -\alpha_2 \\ 1 & -\beta_1 & -\beta_0 & 0 \end{array} \right] \begin{bmatrix} q_t \\ p_t \\ 1 \\ x_t \end{bmatrix} = \begin{bmatrix} \epsilon_t^d \\ \epsilon_t^s \\ \epsilon_t' \end{bmatrix}$$

Entonces

$$\tilde{A}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \tilde{A}_2 = \begin{bmatrix} 1 & -\alpha_2 \\ 1 & 0 \end{bmatrix}$$

Demanda $\tilde{A}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ no cumple condición de orden, por lo tanto es no-identificada.

Oferta $\tilde{A}_2 = \begin{bmatrix} 1 & -\alpha_2 \\ 1 & 0 \end{bmatrix}$ cumple condición de rango si y solo si $\alpha_2 \neq 0$.

En conclusión, podemos estimar la oferta siempre y cuando la demanda efectivamente dependa de x_t . La demanda no puede ser estimada.

Ejemplo 9-4: Identificación de un modelo de demanda agregada

En el ejemplo 9.1 encontramos esta forma estructural para el modelo de oferta y demanda

$$Az_t = \left[\begin{array}{ccc|ccc} 1 & 0 & -\alpha_1 & -\alpha_0 & 0 & 0 \\ 0 & 1 & -\beta_1 & -\beta_0 & \beta_1 & 0 \\ -1 & -1 & 1 & 0 & 0 & -1 \end{array} \right] \begin{bmatrix} C_t \\ I_t \\ Y_t \\ 1 \\ Y_{t-1} \\ G_t \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ 0 \\ z_t \end{bmatrix}$$

En este caso las matrices de restricciones de las ecuaciones 1 y 2 son idénticas:

$$\tilde{A}_1 = \tilde{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \beta_1 & 0 \\ -1 & -1 & 0 & -1 \end{bmatrix}$$

Ec. consumo • Tenemos $\text{rango}[\tilde{A}_1] = 3 \Rightarrow$ identificada.

• \tilde{A}_1 tiene 4 columnas pero 3 filas \Rightarrow sobre-identificada.

Ec. inversión • Dado que $\tilde{A}_2 = \tilde{A}_1$, sabemos que esta ecuación también está sobre-identificada.

Ec. ingreso • Es una identidad \Rightarrow no hay nada que estimar.

El modelo recursivo:

Un sistema de ecuaciones no simultáneas

$$y_1 = x'\beta_1 + \epsilon_1$$

$$y_2 = x'\beta_2 + \gamma_{12}y_1 + \epsilon_2$$

⋮

$$y_M = x'\beta_M + \gamma_{1M}y_1 + \cdots + \gamma_{M-1,M}y_{M-1} + \epsilon_M$$



$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ -\gamma_{12} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_{1M} & -\gamma_{2M} & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} + \begin{bmatrix} -\beta_1 \\ -\beta_2 \\ \vdots \\ -\beta_M \end{bmatrix} \begin{bmatrix} x' \\ \textcolor{teal}{x_t} \\ \textcolor{teal}{B'} \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix}$$

Γ' es triangular

El modelo recursivo:*El término de error*

$$\begin{aligned} \mathbb{E} \epsilon_j &= 0 \\ \mathbb{E} \epsilon_j^2 &= \sigma_j^2 \\ \mathbb{E} \epsilon_i \epsilon_j &= 0 \quad (i \neq j) \end{aligned}$$



$$\mathbb{E}[\epsilon_t \epsilon_t'] = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_M^2 \end{bmatrix}$$

Σ es diagonal

Contando parámetros del modelo recursivo

	parámetros estructurales	parámetros reducidos
Γ	$\frac{M(M-1)}{2}$	
B	KM	Π
Σ	M	Ω
Total:	$\frac{M}{2}(1 + 2K + M)$	Total: $\frac{KM}{2} + \frac{M(M+1)}{2}$

Así, para identificar B, Γ, Σ a partir de Π, Ω tenemos:

$$\frac{M}{2}(1 + 2K + M) - \frac{M}{2}(1 + 2K + M) = \underset{\# \text{ de ecuaciones}}{0} \underset{\# \text{ de incógnitas}}{\text{exceso de parámetros}}$$

es decir, el sistema está exactamente identificado

Ejemplo 9-5: Identificación de un SVAR(1)

- Considere el modelo

$$\begin{aligned} x_t &= \phi y_t + \beta_{11} x_{t-1} + \beta_{21} y_{t-1} + \epsilon_{1t} \\ y_t &= \gamma x_t + \beta_{12} x_{t-1} + \beta_{22} y_{t-1} + \epsilon_{2t} \end{aligned}$$

- o bien, en forma matricial

$$\begin{bmatrix} 1 & -\phi \\ -\gamma & 1 \end{bmatrix}_{\Gamma'} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{bmatrix}_{-B'} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

- Hay 2 ecuaciones pero cada una tiene únicamente una restricción (la normalización), por lo que la condición de orden no se cumple \Rightarrow ninguna ecuación está identificada.
- Para poder estimar este modelo, sería necesario añadir nuevas restricciones.
- Suponga que estamos dispuestos a asumir que x_t no responde a y_t contemporáneamente, es decir $\phi = 0$. Entonces, la forma reducida del modelo sería

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \pi_{11} & \pi_{21} \\ \beta_{12} + \gamma\beta_{11} & \beta_{22} + \gamma\beta_{21} \\ \pi_{12} & \pi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} + \gamma\epsilon_{1t} \end{bmatrix}$$

- Entonces, si lográramos identificar γ los demás parámetros estructurales serían:

$$\begin{aligned} \beta_{11} &= \pi_{11} & \beta_{21} &= \pi_{21} \\ \beta_{12} &= \pi_{12} - \gamma\pi_{11} & \beta_{22} &= \pi_{22} - \gamma\pi_{21} \end{aligned}$$

- Suponga además que estamos dispuestos a asumir que los errores no están correlacionados:

$$\text{Var}[\epsilon_t] = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}_{\Sigma}$$

- Entonces, la varianza de los errores reducidos Ω es:

$$\begin{aligned} \Omega &\equiv \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix}_{\Gamma'^{-1}} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}_{\Sigma} \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}_{\Gamma^{-1}} \\ &= \begin{bmatrix} \sigma_1^2 & \gamma\sigma_1^2 \\ \gamma\sigma_1^2 & \gamma^2\sigma_1^2 + \sigma_2^2 \end{bmatrix} \end{aligned}$$

- Así, conociendo Ω (estimado a partir de la forma reducida), los parámetros estructurales estarían identificados!

$$\sigma_1^2 = \sigma_x^2 \quad \gamma = \frac{\sigma_{xy}}{\sigma_x^2} \quad \sigma_2^2 = \sigma_y^2 - \gamma^2\sigma_x^2$$

- Antes de continuar, observe que

$$\begin{aligned} \Omega &\equiv \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix}_{\Gamma'^{-1}} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}_{\Sigma} \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}_{\Gamma^{-1}} \\ &= \begin{bmatrix} \sigma_1 & 0 \\ \gamma\sigma_1 & \sigma_2 \end{bmatrix}_L \begin{bmatrix} \sigma_1 & \gamma\sigma_1 \\ 0 & \sigma_2 \end{bmatrix}_{L'} \end{aligned}$$

- Es decir, hemos descompuesto Ω como el producto de una matriz triangular inferior y su transpuesta:

$$\Omega = LL'$$

- Esta es la *descomposición de Choleski*. Toda matriz simétrica semi-definida positiva

9.3 El término de error: plim vs \mathbb{E}

- La diagonal de L identifica las desviaciones estándar de los errores estructurales.

Supuestos acerca del término de error: esperanza

Forma estructural	Forma reducida
$\mathbb{E}[\epsilon_t x_t] = 0$	$\mathbb{E}[\nu_t x_t] = (\Gamma^{-1})' 0 = 0$
$\mathbb{E}[\epsilon_t \epsilon_t' x_t] = \Sigma$	$\mathbb{E}[\nu_t \nu_t' x_t] = (\Gamma^{-1})' \Sigma \Gamma^{-1} = \Omega$
$\mathbb{E}[\epsilon_t \epsilon_s' x_t, x_s] = 0$	$\mathbb{E}[\nu_t \nu_s' x_t, x_s] = (\Gamma^{-1})' 0 \Gamma^{-1} = 0$

Nota: plim

Convergencia en probabilidad

- Una secuencia $\{X_n\}$ de variables aleatorias *converge en probabilidad* hacia la variable aleatoria X si para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) = 0$$

- Para denotar que $\{X_n\}$ converge en probabilidad hacia X escribimos

$$X_n \xrightarrow{p} X \quad \text{o bien} \quad \text{plim } X_n = X$$

Propiedades de la convergencia en probabilidad

Sean c una constante, $g()$ una función continua, y X_n, Y_n dos secuencias de variables aleatorias. Entonces:

- $\text{plim } c = c$
- $\text{plim } cX_n = c \text{ plim } X_n$
- $\text{plim } (X_n + Y_n) = \text{plim } X_n + \text{plim } Y_n$
- $\text{plim } (X_n Y_n) = (\text{plim } X_n)(\text{plim } Y_n)$
- $\text{plim } g(X_n) = g(\text{plim } X_n)$

Supuestos acerca del término de error:

límite de probabilidad

Si suponemos que

$$\begin{aligned} \text{plim} \left(\frac{1}{T} E'E \right) &= \Sigma \\ \text{plim} \left(\frac{1}{T} X'X \right) &= Q \\ \text{plim} \left(\frac{1}{T} X'E \right) &= 0 \end{aligned}$$

Tenemos que

$$\text{plim} \left(\frac{1}{T} \begin{bmatrix} Y' \\ X' \\ V' \end{bmatrix} \begin{bmatrix} Y & X & V \end{bmatrix} \right) = \begin{bmatrix} \Pi'Q\Pi + \Omega & \Pi'Q & \Omega \\ Q\Pi & Q & 0' \\ \Omega & 0 & \Omega \end{bmatrix}$$

Supuestos acerca del término de error:

límite de probabilidad

Forma estructural	Forma reducida
$\text{plim} \left(\frac{1}{T} E'E \right) = \Sigma$	$\text{plim} \left(\frac{1}{T} V'V \right) = (\Gamma^{-1})' \Sigma \Gamma^{-1} = \Omega$
$\text{plim} \left(\frac{1}{T} X'X \right) = Q$	$\text{plim} \left(\frac{1}{T} X'X \right) = Q$
$\text{plim} \left(\frac{1}{T} X'E \right) = 0$	$\text{plim} \left(\frac{1}{T} X'V \right) = 0\Gamma^{-1} = 0$

9.4 Estimación

Métodos de estimación

- Se clasifican en métodos indirectos y métodos directos.
- Los directos se clasifican en dos grupos:

Métodos de información limitada

- OLS
- IV
- 2SLS
- GMM
- LIML

Métodos de información completa

- 3SLS
- FIML
- GMM

Consistencia de OLS de la forma reducida

El estimador OLS de las pendientes Π es *consistente*

$$\begin{aligned}
 \text{plim} \hat{\Pi} &= \text{plim} \left[(X'X)^{-1} X'Y \right] \\
 &= \text{plim} \left[\left(\frac{1}{T} X'X \right)^{-1} \frac{1}{T} X'Y \right] \\
 &= \left(\text{plim} \frac{1}{T} X'X \right)^{-1} \text{plim} \left(\frac{1}{T} X'Y \right) \\
 &= Q^{-1} (Q\Pi) \\
 &= \Pi
 \end{aligned}$$

El estimador MCO de la varianza Ω también es *consistente*

$$\begin{aligned}
 \text{plim } \hat{\Omega} &= \text{plim} \left[\frac{1}{T} \hat{V}' \hat{V} \right] \\
 &= \text{plim} \left[\frac{1}{T} \left(Y - X\hat{\Pi} \right)' \left(Y - X\hat{\Pi} \right) \right] \\
 &= \text{plim} \left(\frac{1}{T} Y' Y \right) - \text{plim} \left(\frac{1}{T} Y' X\hat{\Pi} \right) - \text{plim} \left(\hat{\Pi}' \frac{1}{T} X' Y \right) + \text{plim} \left(\hat{\Pi}' \frac{1}{T} X' X\hat{\Pi} \right) \\
 &= \Pi' Q\Pi + \Omega - (\Pi' Q)\Pi - \Pi' (Q\Pi) + \Pi' (Q)\Pi \\
 &= \Omega
 \end{aligned}$$

Mínimos cuadrados indirectos (ILS)

- Es posible estimar Π y Ω consistentemente con OLS, pero estos parámetros no son de interés (excepto para proyectar $Y|X$).
- No obstante, si el sistema está identificado, se puede usar Mínimos Cuadrados Indirectos (ILS), que consiste en estimar $\hat{\Gamma}^{\text{ILS}}$, \hat{B}^{ILS} y $\hat{\Sigma}^{\text{ILS}}$ en función de $\hat{\Pi}^{\text{OLS}}$ y $\hat{\Omega}^{\text{OLS}}$.
- Propiedades:
 - factible pero ineficiente
 - puede haber más de una solución (si está sobre- identificado).

Mínimos cuadrados ordinarios (OLS)

Escribimos la ecuación j como

$$y_j = Y_j \gamma_j + X_j \beta_j + \epsilon_j = [Y_j \quad X_j] \begin{bmatrix} \gamma_j \\ \beta_j \\ \delta_j \end{bmatrix} + \epsilon_j = Z_j \delta_j + \epsilon_j$$

Estimador OLS

$$\begin{aligned}
 \hat{\delta}_j^{\text{OLS}} &= (Z_j' Z_j)^{-1} Z_j' y_j \\
 &= (Z_j' Z_j)^{-1} Z_j' (Z_j \delta_j + \epsilon_j) \\
 &= \delta_j + (Z_j' Z_j)^{-1} Z_j' \epsilon_j \\
 &= \delta_j + \begin{bmatrix} Y_j' Y_j & Y_j' X_j \\ X_j' Y_j & X_j' X_j \end{bmatrix}^{-1} \begin{bmatrix} Y_j' \epsilon_j \\ X_j' \epsilon_j \end{bmatrix}
 \end{aligned}$$

OLS es *inconsistente* porque $\text{plim} \left[\frac{1}{T} Y_j' \epsilon_j \right] \neq 0$ (sesgo de simultaneidad)

Caso particular: Modelo recursivo

En el modelo recursivo donde Γ es triangular y Σ es diagonal,

$$\begin{aligned}
 y_1 &= x' \beta_1 + \epsilon_1 \\
 y_2 &= x' \beta_2 + \gamma_{12} y_1 + \epsilon_2 \\
 &\vdots \\
 y_M &= x' \beta_M + \gamma_{1M} y_1 + \cdots + \gamma_{M-1,M} y_{M-1} + \epsilon_M
 \end{aligned}$$

OLS es consistente y eficiente, porque no hay sesgo de simultaneidad:

$$\begin{aligned}\text{Cov}[y_1, \epsilon_2] &= \text{Cov}[x'\beta_1 + \epsilon_1, \epsilon_2] = 0 \\ \text{Cov}[y_1, \epsilon_3] &= \text{Cov}[x'\beta_1 + \epsilon_1, \epsilon_3] = 0 \\ \text{Cov}[y_2, \epsilon_3] &= \text{Cov}[x'\beta_2 + \gamma_{12}y_1 + \epsilon_2, \epsilon_3] = 0\end{aligned}$$

y así sucesivamente.

Variables instrumentales (IV)

- Para la ecuación $y_j = Y_j\gamma_j + X_j\beta_j + \epsilon_j = Z_j\delta_j + \epsilon_j$ suponemos que existe matriz W_j $T \times (M_j + K_j)$ tal que:

$$\begin{aligned}\text{plim}\left(\frac{1}{T}W'_jZ_j\right) &= \Sigma_{WZ} && (\text{instrumento correlacionado con regresores}) \\ \text{plim}\left(\frac{1}{T}W'_jW_j\right) &= \Sigma_{WW} && (\text{instrumento tiene varianza finita}) \\ \text{plim}\left(\frac{1}{T}W'_j\epsilon_j\right) &= 0 && (\text{instrumento NO correlacionado con errores})\end{aligned}$$

Estimador IV

$$\begin{aligned}\hat{\delta}_j^{\text{IV}} &= (W'_jZ_j)^{-1} W'_j y_j \\ &= \delta_j + \left(\frac{1}{T}W'_jZ_j\right)^{-1} \left(\frac{1}{T}W'_j\epsilon_j\right)\end{aligned}$$

IV es consistente porque $\text{plim}\left[\frac{1}{T}W'_j\epsilon_j\right] = 0$

Mínimos cuadrados en dos etapas (2SLS)

Consiste en usar como instrumentos para Y_j los valores ajustados por la regresión de Y_j en *todas* las X 's del sistema.

Etapa 1 Ajustar Y_j por OLS usando todas las X

$$\hat{Y}'_j^{\text{OLS}} = X\hat{\Pi}'_j^{\text{OLS}} = X(X'X)^{-1}X'Y_j$$

Etapa 2 Usar IV con $W = [\hat{Y}'_j^{\text{OLS}} \quad X]$:

Estimador 2SLS

$$\hat{\delta}_j^{\text{2SLS}} = \begin{bmatrix} \hat{Y}'_j^{\text{OLS}} Y_j & \hat{Y}'_j^{\text{OLS}} X_j \\ X'_j Y_j & X'_j X_j \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}'_j^{\text{OLS}} y_j \\ X'_j y_j \end{bmatrix}$$

Si no hay autocorrelación ni heteroscedasticidad entonces 2SLS es el estimador IV más eficiente usando sólo la información de X

Ejemplo 9-6: El Modelo Klein I

 TableF10-3.txt

 klein-model.do

En 1950 Klein estimó este modelo (conocido ahora como el modelo I de Klein) con datos

anuales de 1921 a 1941^a:

$$\begin{aligned}
 C_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (Wp_t + Wg_t) + \epsilon_{1t} \\
 &\text{consumo} \\
 I_t &= \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \epsilon_{2t} \\
 &\text{inversión} \\
 Wp_t &= \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + \epsilon_{3t} \\
 &\text{salarios} \\
 &\text{privados} \\
 X_t &= C_t + I_t + G_t \\
 &\text{producto} \\
 P_t &= X_t - T_t - Wp_t \\
 &\text{utilidades} \\
 K_t &= K_{t-1} + I_{t-1} \\
 &\text{capital}
 \end{aligned}$$

- Las variables exógenas son:
 - G_t = gasto (no salarial) del gobierno
 - T_t = impuestos indirectos a las empresas + exportaciones netas
 - Wg_t = gastos salarial del gobierno
 - A_t = tendencia, años desde 1931
- Hay tres variables predeterminadas: los rezagos del stock de capital, utilidades privadas, y demanda total.
- El modelo contiene 3 ecuaciones de comportamiento, una condición de equilibrio, y dos identidades contables.

El estimador 2SLS del sistema es (estadístico p en paréntesis):

$$\begin{aligned}
 C_t &= 16.55 + 0.02 P_t + 0.22 P_{t-1} + 0.81 (Wp_t + Wg_t) + \epsilon_{1t} \\
 &\quad (0.00) \quad (0.88) \quad (0.04) \quad (0.00) \\
 I_t &= 20.28 + 0.15 P_t + 0.62 P_{t-1} - 0.16 K_{t-1} + \epsilon_{2t} \\
 &\quad (0.01) \quad (0.39) \quad (0.00) \quad (0.00) \\
 Wp_t &= 1.50 + 0.44 X_t + 0.15 X_{t-1} + 0.13 A_t + \epsilon_{3t} \\
 &\quad (0.19) \quad (0.00) \quad (0.00) \quad (0.00)
 \end{aligned}$$

^aBasado en Greene (2012, pp332-333)

Ejemplo 9-7: El modelo de la telaraña



En el modelo de la telaraña para los mercados de maíz y de trigo

$$\begin{array}{lllll}
 q_{mt}^s = 200 & +0.5p_{m,t-1} & -0.2p_{w,t-1} & +\epsilon_{mt}^s & (\text{oferta de maíz}) \\
 q_{wt}^s = 80 & -0.2p_{m,t-1} & +0.4p_{w,t-1} & +\epsilon_{wt}^s & (\text{oferta de trigo}) \\
 q_{mt}^d = 100 & -0.2p_{mt} & +1.2p_{wt} & +\epsilon_{mt}^d & (\text{demanda de maíz}) \\
 q_{wt}^d = 50 & +1.1p_{mt} & -0.4p_{wt} & +\epsilon_{wt}^d & (\text{demanda de trigo})
 \end{array}$$

- Los agricultores pueden sembrar maíz o trigo, pero tardan un período en producirlo.

- Los consumidores están dispuestos a sustituir el consumo de maíz y trigo en respuesta a los precios que encuentran.

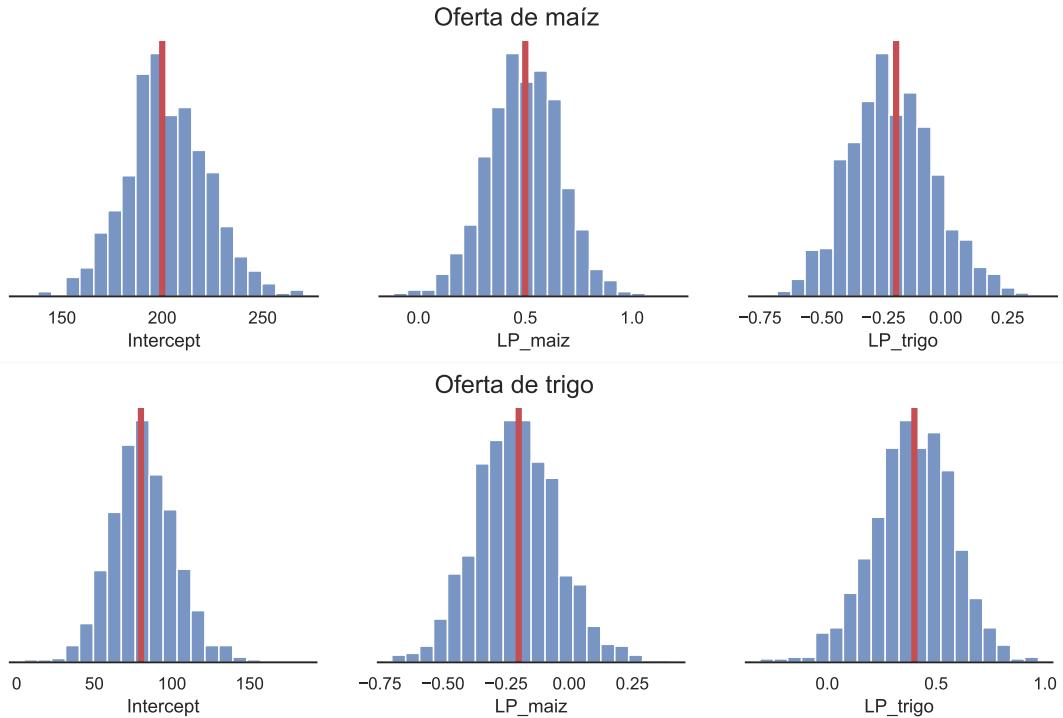
Este modelo puede escribirse

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0.2 & -1.2 \\ 0 & 1 & -1.1 & 0.4 \end{bmatrix} \begin{bmatrix} q_{mt} \\ q_{wt} \\ p_{mt} \\ p_{wt} \end{bmatrix} = \begin{bmatrix} 200 & 0.5 & -0.2 \\ 80 & -0.2 & 0.4 \\ 100 & 0 & 0 \\ 50 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ p_{m,t-1} \\ p_{w,t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{mt}^s \\ \epsilon_{wt}^s \\ \epsilon_{mt}^d \\ \epsilon_{wt}^d \end{bmatrix}$$

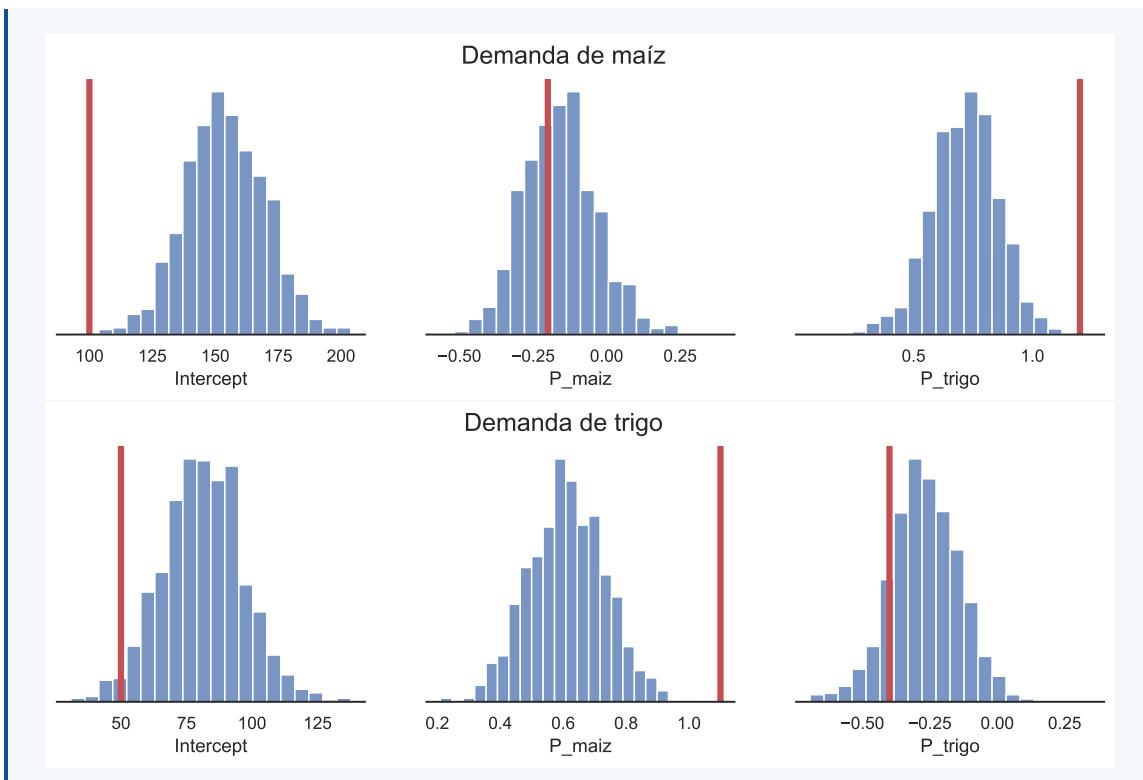
$\Gamma \qquad \qquad \qquad B \qquad \qquad \qquad \epsilon_t$

- Vemos que las ecuaciones de oferta no son simultáneas, pero las de demanda sí lo son.
- Para ilustrar el sesgo de simultaneidad del estimador OLS, y que el estimador 2SLS sí es insesgado, realizamos experimentos de Monte Carlo.
- En particular estimamos 1000 veces este modelo a partir de series simuladas de 24 observaciones.
- En las figuras que siguen vemos la distribución obtenida para cada parámetro y lo comparamos con el verdadero valor poblacional.

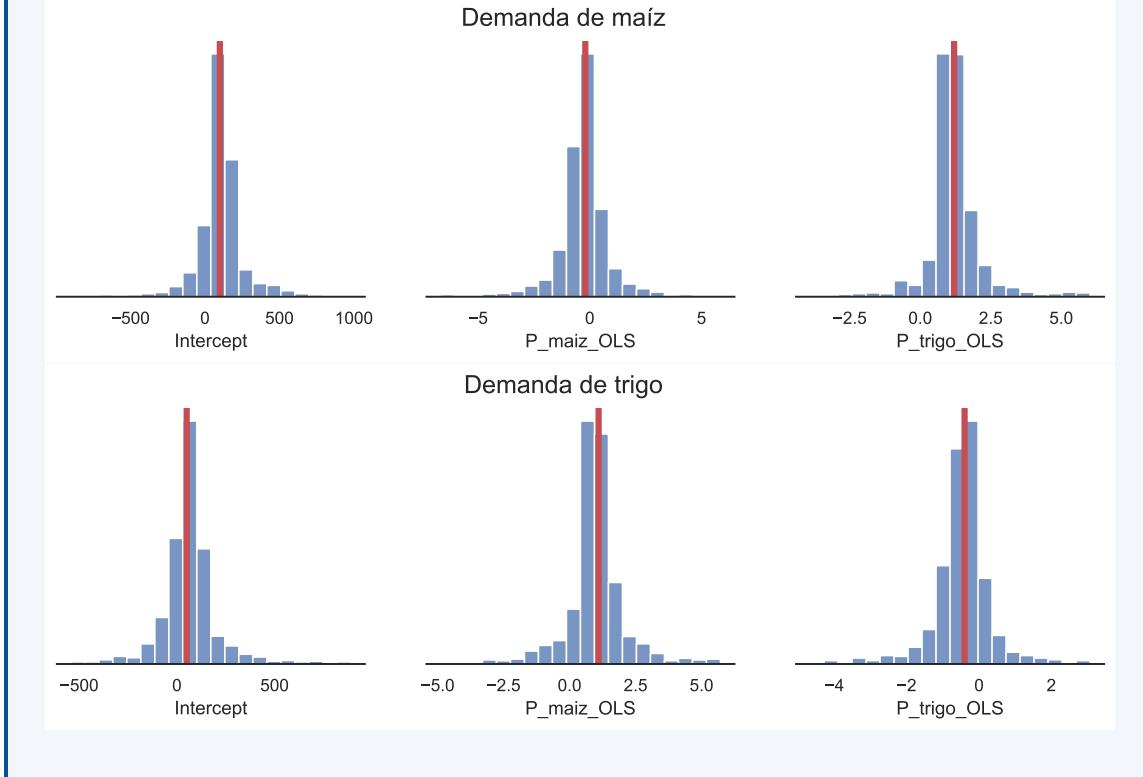
Las ecuaciones de oferta no son simultáneas, por lo que pueden ser estimadas por OLS.



Las ecuaciones de demanda sí son simultáneas, por lo que la estimación OLS estaría sesgada.



Por ello, estimamos el sistema por 2SLS.



Referencias del capítulo 9

Greene, William H. (2012). *Econometric Analysis*. 7^a ed. Prentice Hall. ISBN: 978-0-13-139538-1.

Capítulo 10

Vectores Auto-Regresivos (VAR)

En este capítulo

10.1 Representaciones alternativas de un VAR	168
10.2 La función de impulso respuesta	172
10.3 Estacionariedad	173
10.4 Estimación	176
10.5 Causalidad de Granger	180
10.6 Pronóstico y la descomposición de varianza de su error	180

Las tareas del macroeconometrista ¹

En general, los macroeconometristas hacemos cuatro cosas:

1. describimos y resumimos datos macroeconómicos
2. realizamos pronósticos macroeconómicos
3. cuantificamos lo que sabemos (y lo que no sabemos) acerca de la verdadera estructura de la macroeconomía
4. aconsejamos a quienes toman decisiones de política (y a veces tomamos esas decisiones).

Un poco de historia

- En los 1970s, estas cuatro tareas se realizaban usando una variedad de técnicas.
- Pero después del caos macroeconómico de los 1970s, ninguna de esas técnicas parecía apropiada.
- En 1980, Sims (1980, pp.15) afirma que

Debe ser factible estimar modelos macro de gran escala como formas reducidas sin restricciones, tratando todas las variables como endógenas.

¹Esta introducción está basada en Stock y Watson (2001)

La propuesta de Sims

- En 1980, Sims propone una nueva técnica macroeconómica: el vector autorregresivo (VAR)
- Un VAR es un modelo lineal de n ecuaciones y n variables, en el cual cada variable es explicada por sus propios rezagos, más los valores actuales y rezagados de las otras $n - 1$ variables.
- Este sencillo esquema provee una manera sistemática de capturar la gran dinámica en series de tiempo múltiples, y las herramientas estadísticas que vienen con los VAR son fáciles de usar y de interpretar.

10.1 Representaciones alternativas de un VAR

Las series de tiempo

- En lo que sigue, asumimos que hay n series de tiempo.
- La observación t se denota y_t , corresponde a un vector columna con n datos (uno por cada serie)

	A	B	C	D	E	F	G
1	fecha	variable1	variable2	variable3			
2	1	-1.4	0.3	2			
3	2	0.6	-1.0	4			
4	3	-0.1	0.9	1			
5	4	0.5	4.1	4			
6	5	1.6	2.6	5			
7	6	1.5	0.9	4	y_t		
8	7	-1.1	-0.4	5			
9	8	0.5	4.8	4	y_{t+2}		
10							

Distintos “sabores” de VAR

- En general, hay tres variedades de VAR:
 1. VAR en forma reducida
 2. VAR recursivo
 3. VAR estructural
- Se distinguen por cómo se presentan las relaciones *contemporáneas* entre las variables.

El proceso VAR(p) en forma reducida

- El VAR reducido expresa cada variable como función lineal de los rezagos de todas las variables en el sistema y de un término de error sin correlación serial.
- Si hay p rezagos, el VAR reducido se representa

VAR(p)

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

donde Φ_j es la matriz $n \times n$ de coeficientes del rezago j .

- El término de error cumple:

$$\mathbb{E} \epsilon_t = 0 \quad \text{Var } \epsilon_t = \Omega \quad \text{Cov}[\epsilon_t, \epsilon_s] = 0 \text{ if } t \neq s$$

- *No hay relación contemporánea entre las variables.*

El proceso VAR(p) estructural

- El VAR estructural es similar al reducido, pero *las variables tienen relación contemporánea*.
- Para determinarla, se usa teoría económica.
- Se requiere de *supuestos de identificación* que permitan interpretar correlaciones como causalidad.
- Se representa como

$$\Gamma_0 y_t = d + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t$$

donde Γ_j es la matriz $n \times n$ de coeficientes del rezago j .

- El término de error cumple:

$$\mathbb{E} \varepsilon_t = 0 \quad \text{Cov}[\varepsilon_{it}, \varepsilon_{js}] = \begin{cases} \sigma_{ij}^2, & \text{si } t = s \\ 0, & \text{de lo contrario.} \end{cases}$$

El proceso VAR(p) en forma recursiva

- El VAR recursivo es un caso particular del VAR estructural, donde las variables tienen relación contemporánea de manera recursiva:
 - y_{1t} es “exógena”
 - y_{2t} depende sólo de y_{1t}
 - y_{3t} depende de y_{1t} y de y_{2t}
 - y_{nt} depende de todas las demás.
- Note que el ordenamiento de las variables es importante, habiendo $n!$ formas de ordenarlas. (Por ejemplo, en un VAR de cinco variables, hay $5! = 120$ ordenamientos).
- Al igual que el VAR estructural, se representa como

$$\Gamma_0 y_t = d + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t$$

pero con la restricción de que Γ_0 es uni-triangular inferior.

Ejemplo 10-1: Distintas versiones de un VAR

Forma reducida

$$\begin{aligned} y_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\ m_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\ z_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z \end{aligned}$$

Forma estructural

$$\begin{aligned} y_t + \beta_1 m_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\ m_t + \beta_2 z_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\ z_t + \beta_3 y_t + \beta_4 m_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z \end{aligned}$$

Forma recursiva

$$\begin{aligned} y_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\ \gamma_1 y_t + m_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\ \gamma_2 y_t + \gamma_3 m_t + z_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z \end{aligned}$$

Pasando de un VAR estructural (o recursivo) a uno reducido

- Si la matriz Γ_0 es invertible (lo que está garantizado en el VAR recursivo), entonces se puede obtener el VAR reducido así:

$$\begin{aligned}\Gamma_0 y_t &= d + \Gamma_1 y_{t-1} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t \\ y_t &= \Gamma_0^{-1} d + \Gamma_0^{-1} \Gamma_1 y_{t-1} + \cdots + \Gamma_0^{-1} \Gamma_p y_{t-p} + \Gamma_0^{-1} \varepsilon_t \\ &= c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t\end{aligned}$$

- Las matrices de covarianza cumplen

$$\Omega = \Gamma_0^{-1} \Sigma \Gamma_0^{-1}$$

La descomposición de Cholesky

- Recuerde que

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \Rightarrow \Sigma^{1/2} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}$$

- Así

$$\Omega = [\Gamma_0^{-1} \Sigma^{1/2}] [\Gamma_0^{-1} \Sigma^{1/2}]'$$

- En el caso de un VAR recursivo, Γ_0 es tringular inferior, y por tanto $\Gamma_0^{-1} \Sigma^{1/2}$ también lo es.
- Así, hemos obtenido la descomposición de Cholesky de Ω .

Ejemplo 10-2: Descomposición de Cholesky

$$\Omega = PP' = [\Gamma_0^{-1} \Sigma^{1/2}] [\Gamma_0^{-1} \Sigma^{1/2}]' = \Gamma_0^{-1} \Sigma \Gamma_0^{-1}$$

Si $y'_t = [m_t \quad r_t \quad k_t]$ y la matriz de covarianza reducida es

$$\begin{aligned}\Omega &= \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 4.25 & 2.5 \\ -1 & 2.5 & 12.25 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ -1 & 1.5 & 3 \end{bmatrix}}_P \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 2 & 1.5 \\ 0 & 0 & 3 \end{bmatrix}}_{P'} \\ &= \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -1 & 0.75 & 1 \end{bmatrix}}_{\Gamma_0^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix}}_{\Gamma'^{-1}_0}\end{aligned}$$

Por lo que

$$\Gamma_0 = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ 1.375 & -0.75 & 1 \end{bmatrix}$$

Entonces $\sigma_r^2 = 4$ y un aumento de una unidad en m_t provoca una *disminución contemporánea* de 1.375 en k_t .

VAR con series como desviación de la media

- Si el VAR es estacionario, su media $\mathbb{E} y_t \equiv \mu_t$ es constante
- Así,

$$\begin{array}{rcl} y_t = \Phi_1 y_{t-1} & + \cdots + \Phi_p y_{t-p} & + c + \epsilon_t \\ \mu = \Phi_1 \mu & + \cdots + \Phi_p \mu & + c + \mathbb{E} \epsilon_t \\ \hline y_t - \mu = \Phi_1 (y_{t-1} - \mu) & + \cdots + \Phi_p (y_{t-p} - \mu) & + \epsilon_t \end{array}$$

- La media del proceso es

$$\mu = (I - \Phi_1 - \Phi_2 - \cdots - \Phi_p)^{-1} c$$

- Definimos $\hat{y} \equiv y_t - \mu$. El VAR(p) con series como desviación de la media es

$$\hat{y}_t = \Phi_1 \hat{y}_{t-1} + \cdots + \Phi_p \hat{y}_{t-p} + \epsilon_t$$

VAR(p) escrito como VAR(1)

$$\begin{bmatrix} \hat{y}_t \\ \hat{y}_{t-1} \\ \hat{y}_{t-2} \\ \vdots \\ \hat{y}_{t-p+1} \\ \hat{Y}_t \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & 0 & \cdots & 0 & 0 \\ 0 & I & 0 & \cdots & 0 & 0 \\ 0 & 0 & I & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I & 0 \\ \Phi & & & & & \end{bmatrix} \begin{bmatrix} \hat{y}_{t-1} \\ \hat{y}_{t-2} \\ \hat{y}_{t-3} \\ \vdots \\ \hat{y}_{t-p} \\ \hat{Y}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \\ \xi_t \end{bmatrix}$$

así

$$\hat{Y}_t = \Phi \hat{Y}_{t-1} + \xi_t$$

Ejemplo 10-3: VAR(2) a VAR(1)

El VAR(2)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix}$$

se escribe como VAR(1)

$$\begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \left[\begin{array}{cc|cc} .5 & .1 & 0 & 0 \\ .4 & .5 & .25 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array} \right] \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \\ 0 \\ 0 \end{bmatrix}$$

Dinámica de un proceso VAR(1)

- A menudo necesitamos iterar en la fórmula del AR(1) para analizar su dinámica.
- En esos casos, es útil observar que:

$$(1 + \Phi L + \Phi^2 L^2 + \cdots + \Phi^s L^s)(I - \Phi L) = (I - \Phi^{s+1} L^{s+1})$$

- Así, si queremos expresar \hat{Y}_t en términos de \hat{Y}_{t-s-1}

$$\begin{aligned} (I - \Phi L) \hat{Y}_t &= \xi_t \\ (I - \Phi^{s+1} L^{s+1}) \hat{Y}_t &= (1 + \Phi L + \Phi^2 L^2 + \cdots + \Phi^s L^s) \xi_t \\ \hat{Y}_t - \Phi^{s+1} \hat{Y}_{t-s-1} &= \xi_t + \Phi \xi_{t-1} + \cdots + \Phi^s \xi_{t-s} \end{aligned}$$

Pasado versus futuro

Pasado Cuando queremos analizar \hat{Y}_t en función de shocks pasados, utilizamos

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \cdots + \Phi^s \xi_{t-s} + \Phi^{s+1} \hat{Y}_{t-s-1}$$

Futuro Para analizar el efecto de nuevos shocks sobre futuros valores de \hat{Y} , aplicamos L^{-s} a la última ecuación:

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

VAR(1) escrito como VMA(∞)

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \cdots + \Phi^s \xi_{t-s} + \Phi^{s+1} \hat{Y}_{t-s-1}$$

- Si los eigenvalores de Φ están en el círculo unitario el límite $s \rightarrow \infty$ converge a

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \Phi^2 \xi_{t-2} + \dots$$

10.2 La función de impulso respuesta

Impulso respuesta

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

- La función de *impulso-respuesta* mide la respuesta observada en la variable m -ésima s períodos después ($\hat{Y}_{t+s,m}$) de que se presenta un impulso en la k -ésima variable ($\xi_{t,k}$)
- Viene dada por

$$\frac{\partial \hat{Y}_{t+s,m}}{\partial \xi_{t,k}} = (\Phi^s)_{km}$$

es decir, por el elemento en la fila k , columna m , de la matriz Φ elevada al número de períodos s .

Interpretando la función de impulso-respuesta

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

- Suponga que
 - el sistema estaba en equilibrio en $t-1$, es decir $\hat{Y}_{t-1} = 0$
 - hay un shock $v' = [v_1 \dots v_n]$ a las variables en t , $\xi_t = v$
 - el shock es transitorio: $0 = \xi_{t+1} = \xi_{t+2} = \dots$
- En este caso, la desviación del sistema respecto a su equilibrio s períodos después del shock es

$$\hat{Y}_{t+s} = \Phi^s v$$

Impulso respuesta**(cont'n)**

$$\Omega = PP' = [\Gamma_0^{-1}\Sigma^{1/2}] [\Gamma_0^{-1}\Sigma^{1/2}]' = \Gamma_0^{-1}\Sigma\Gamma_0^{-1}$$

- En la práctica, estamos interesados en shocks ε_t a las ecuaciones estructurales, en vez de a las ecuaciones reducidas ϵ_t . Esto para tomar en cuenta los efectos contemporáneos del shock.
- Para calcular las respuestas, nos valemos de $\epsilon_t = \Gamma_0^{-1}\varepsilon_t$ y de la descomposición de Cholesky de la covarianza de los errores reducidos Ω :

Tamaño del impulso	Impulso estructural	Impulso reducido	Respuesta
unitarios	$\varepsilon_t = I$	$v = \Gamma_0^{-1}$	$\Phi^s\Gamma_0^{-1}$
1 st. dev.	$\varepsilon_t = \Sigma^{1/2}$	$v = \Gamma_0^{-1}\Sigma^{1/2}$	$\Phi^s\Gamma_0^{-1}\Sigma^{1/2}$

Ejemplo 10-4: Impulso respuesta y descomposición de Cholesky

Siguiendo con el ejemplo 10.1, si $y'_t = [m_t \ r_t \ k_t]$ y la matriz de covarianza reducida es

$$\begin{aligned} \Omega &= \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 4.25 & 2.5 \\ -1 & 2.5 & 12.25 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ -1 & 1.5 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 2 & 1.5 \\ 0 & 0 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -1 & 0.75 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix} \\ &\quad \Gamma_0^{-1} \quad \Sigma \quad \Gamma_0'^{-1} \end{aligned}$$

La respuesta del sistema a un shock en k_t se calcula a partir de ...

- $v' = [0 \ 2 \ 1.5]$ si el shock es de una desviación estándar.
- $v' = [0 \ 1 \ 0.75]$ si el shock es unitario.

10.3 Estacionariedad

Estacionariedad

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi\xi_{t+s-1} + \dots + \Phi^s\xi_t + \Phi^{s+1}\hat{Y}_{t-1}$$

- Recuerde que la respuesta del VAR s períodos después a un impulso ξ_t es $\Phi^s\xi_t$.
- Si el VAR es estacionario, toda respuesta a cualquier impulso en t debe ser transitoria: $\lim_{s \rightarrow \infty} \Phi^s\xi_t = 0$. De lo contrario, la media μ del proceso no sería constante.
- Es decir, el VAR es estacionario si y sólo si

$$\lim_{s \rightarrow \infty} \Phi^s = 0$$

Nota: Descomposición espectral de una matriz

Descomposición espectral de una matriz

Si los eigenvectores de la matriz cuadrada A son linealmente independientes, entonces

Descomposición espectral

$$A = C\Lambda C^{-1}$$

donde Λ es la matriz diagonal formada por los eigenvalores de A :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

y las columnas de C son los correspondientes eigenvectores de A .

Potencia de una matriz

Si A tiene la descomposición espectral $A = C\Lambda C^{-1}$ es fácil calcular su t -ésima potencia:

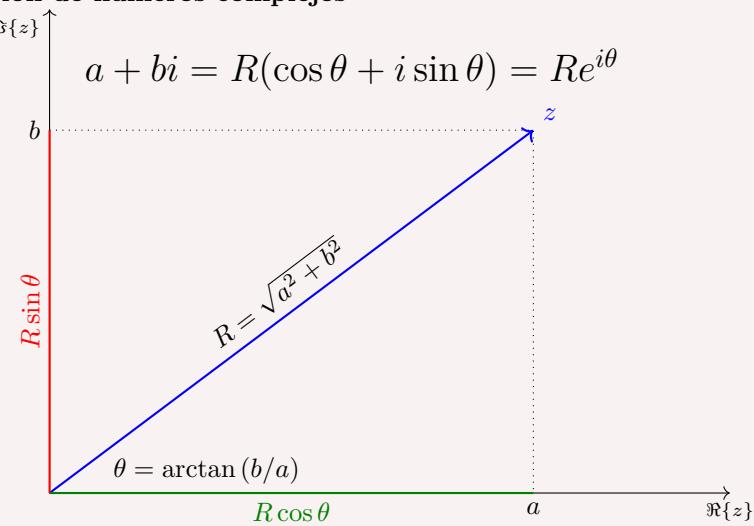
$$A^t = C\Lambda^t C^{-1}$$

ya que

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & 0 & \dots & 0 \\ 0 & \lambda_2^t & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n^t \end{bmatrix}$$

Nota: Números complejos

Representación de números complejos



Multiplicación de números complejos

- Si $z = Re^{i\theta}$ y $w = Se^{i\varphi}$, entonces su producto es

$$zw = RSe^{i(\theta+\varphi)}$$

- Así, si elevamos z a la n -ésima potencia:

$$z^n = (Re^{i\theta})^n = R^n e^{in\theta}$$

- Es decir

$$\lim_{n \rightarrow \infty} z^n = 0 \Leftrightarrow |R| < 1$$

Ejemplo de potencia de números complejo

Condiciones para la estacionariedad

- Y es estacionario si y solo si todos los eigenvalores de Φ están dentro del círculo unitario.
- Los eigenvalores λ de Φ satisfacen:

$$|I\lambda^p - \Phi_1\lambda^{p-1} - \cdots - \Phi_p| = 0$$

Ejemplo 10-5: Estacionariedad de un VAR(2)

Para el VAR(2) del ejemplo 10.1:

$$\begin{aligned} 0 &= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \lambda^2 - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \lambda - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} \lambda^2 - .5\lambda & -.1\lambda \\ -.4\lambda - .25 & \lambda^2 - .5\lambda \end{bmatrix} \right| \\ &= (\lambda^2 - .5\lambda)(\lambda^2 - .5\lambda) - (-.4\lambda - .25)(-.1\lambda) \\ &= \lambda^4 - \lambda^3 + 0.21\lambda^2 - 0.025\lambda \end{aligned}$$

Las raíces de este polinomio son $\lambda_1 = 0$, $\lambda_2 = 0.7693$, $\lambda_3 = 0.1154 + 0.1385i$, y $\lambda_4 = 0.1154 - 0.1385i$, todas ellas dentro del círculo unitario. Por lo tanto, el VAR es estacionario.

Ejemplo 10-6: Dinámica de un VAR

- Hasta el momento hemos estudiado
 - cómo determinar si un VAR es estacionario
 - la función de impulso respuesta
- En el cuaderno de Jupyter 02 `Simulacion de un VAR.ipynb` se introduce un ejemplo para analizar un VAR.
- En el cuaderno de Jupyter 03 `VAR(1) clase.ipynb` se presentan 5 modelos VAR reducidos. Acá definimos una *clase* para representar un VAR.

10.4 Estimación

Especificación de un VAR

- Para especificar un VAR, hay dos cosas que escoger:
 - Las n series que conforman el VAR
 - El número p de rezagos a ser incluidos en el modelo
- En contraste con la metodología de Box-Jenkins, el modelo VAR no es parsimonioso \Rightarrow un VAR está sobreparametrizado:
 - hay $n + pn^2$ parámetros.
 - muchos de sus coeficientes no son significativos
 - regresores posiblemente son altamente colineales, estadísticos t no confiables.
- No obstante, en un VAR los coeficientes individuales no son de interés. Interesa determinar la dinámica del modelo (impulso respuesta) y la causalidad (Granger).

Estimación de un VAR

- El VAR reducido es un sistema de ecuaciones SUR, donde todas las ecuaciones tienen los mismos regresores.
- Por lo tanto, la estimación OLS ecuación por ecuación es insesgada, eficiente y consistente.
- La matriz de covarianza de los errores se estima a partir de los residuos

$$\hat{\Omega} = \frac{1}{T - np - 1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'$$

Estimación de un VAR no estacionario

Si las series son I(1), ¿hay que diferenciarlas para estimar el VAR? Hay un debate respecto:

NO: Algunos aconsejan no diferenciar las series, porque se pierde información, alegando que la meta del VAR es determinar relaciones entre variables, no los parámetros mismos.

SI: Otros advierten que si las series son integradas, el VAR en niveles no es consistente con el proceso generador de datos. En este caso se aconseja estudiar la cointegración de las series \Rightarrow VECM

Ejemplo 10-7: Propiedades de la estimación OLS de un VAR

 `VAR_insesgado.ipynb`

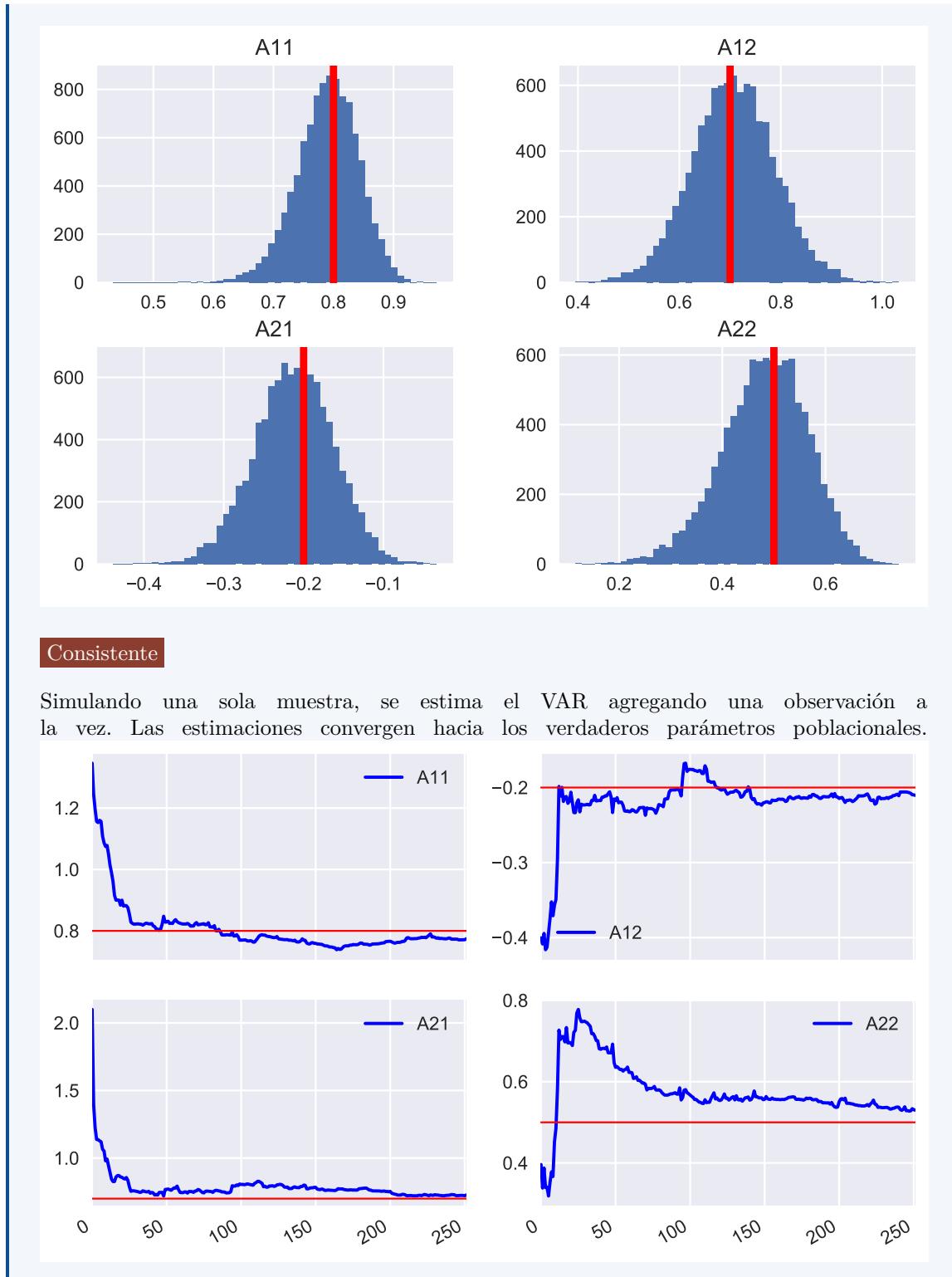
- En el cuaderno de Jupyter `VAR_insesgado.ipynb` se presenta un ejercicio de simulación para ilustrar que el estimador OLS de un VAR es insesgado y consistente.
- Se simula el VAR(1)

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix}_{Y_t} = \begin{pmatrix} 0.8 & 0.7 \\ -0.2 & 0.5 \end{pmatrix}_{\Phi} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}_{Y_{t-1}} + \begin{pmatrix} \epsilon_{x,t} \\ \epsilon_{y,t} \end{pmatrix}_{\epsilon_t}$$

$$\Omega = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$$

Insesgado

Con un experimento de Monte Carlo de las 10.000 repeticiones obtenemos esta distribución de las estimaciones.



Escogiendo el óptimo número de rezagos del VAR

- La selección del número de rezagos p es crítica.

- Si p es demasiado pequeño, el modelo está mal especificado
- Si p es demasiado grande, se desperdician grados de libertad y se introduce colinealidad.
- Una opción: dejar que p difiera para cada ecuación y variable \Rightarrow el VAR no podría estimarse con OLS.
- Por tanto, se determina un sólo p para todas las variables y ecuaciones, usando dos enfoques distintos
 - Descartando los últimos rezagos si no son significativos
 - Utilizando un criterio de selección que contraponga el ajuste a los grados de libertad.

Escogiendo el p óptimo con pruebas de significancia

- Suponga que queremos decidir entre p y q número de rezagos, con $q < p$.
- Obtenemos $\hat{\Omega}$ para ambos modelos, *usando la misma muestra*.
- Comprobamos la significancia conjunta de los rezagos $q+1, \dots, p$ con la prueba de razón de verosimilitud

$$(T - 1 - pn) \left(\ln |\hat{\Omega}_q| - \ln |\hat{\Omega}_p| \right) \xrightarrow{asym} \chi^2(n^2(p-q))$$
- Si no se rechaza la hipótesis nula, puede escogerse q rezagos.
- Limitaciones:
 - Resultado asintótico, pero muestras usualmente pequeñas
 - Resultado sensible a escogencia de p y q .

Escogiendo el p óptimo con criterios de selección

- En la práctica, el número óptimo de rezagos se escoge minimizando el valor de AIC o de SBC:

$$\begin{aligned} \text{AIC} &= T \ln |\hat{\Omega}| + 2N \\ \text{SBC} &= T \ln |\hat{\Omega}| + N \ln T \end{aligned}$$

donde $N = n^2p + n$ es el número total de parámetros estimados

- Rezagos adicionales reducen $\hat{\Omega}$, pero incrementando N

Usando un VAR para describir los datos

- La práctica usual en el análisis VAR es reportar
 - pruebas de causalidad de Granger
 - funciones de impulso respuesta
 - descomposición de varianza del error de pronóstico
- Usualmente no se reportan los coeficientes estimados ni el R^2 , porque no son tan informativos.

10.5 Causalidad de Granger

Causalidad de Granger

- Las pruebas de causalidad de Granger determinan si los valores rezagados de una variable ayudan a predecir otra variable.
- Si los rezagos de A no ayudan a predecir B , entonces coeficientes de los rezagos de A en la ecuación de B son ceros.
- Si alguna variable no es causada en el sentido de Granger por ninguna otra, *no implica que la variable sea exógena*. Bien podría haber causalidad contemporánea.

Ejemplo 10-8: Causalidad de Granger

En el modelo VAR(2) con variables y_t, m_t :

$$\begin{aligned} y_t &= c^y + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \beta_{11}y_{t-2} + \beta_{12}m_{t-2} + u_t^y \\ m_t &= c^m + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \beta_{21}y_{t-2} + \beta_{22}m_{t-2} + u_t^m \end{aligned}$$

las pruebas de causalidad de Granger son:

Hipótesis nula	Restricciones
m no causa y	$\alpha_{12} = \beta_{12} = 0$
y no causa m	$\alpha_{21} = \beta_{21} = 0$

Si se rechaza, por ejemplo, la primera hipótesis, decimos que m *causa y en el sentido de Granger*.

10.6 Pronóstico y la descomposición de varianza de su error

Pronóstico

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi\xi_{t+s-1} + \cdots + \Phi^{s-1}\xi_{t+1} + \Phi^s\hat{Y}_t$$

- Suponga que Φ ha sido estimado con datos hasta $t = T$.
- El mejor pronóstico del sistema s períodos adelante es

$$\mathbb{E} [\hat{Y}_{T+s} | \hat{Y}_T] = \Phi^s\hat{Y}_T$$

- El error de pronóstico es

$$\hat{Y}_{T+s} - \mathbb{E} [\hat{Y}_{T+s} | \hat{Y}_T] = \xi_{t+s} + \Phi\xi_{t+s-1} + \cdots + \Phi^{s-1}\xi_{t+1}$$

- y su varianza (MSPE) es

$$\text{Var} [\hat{Y}_{T+s} | \hat{Y}_T] = \Omega + \Phi\Omega\Phi' + \cdots + \Phi^{s-1}\Omega\Phi'^{s-1}$$

Pronósticos de largo plazo con VAR

$$\mathbb{E} [\hat{Y}_{T+s} | \hat{Y}_T] = \Phi^s \hat{Y}_T$$

- Recuerde que:
 - $\hat{y} \equiv y_t - \mu$
 - si el VAR es estacionario, $\lim_{s \rightarrow \infty} \Phi^s = 0$.
- Entonces el pronóstico de largo plazo

$$\lim_{s \rightarrow \infty} \mathbb{E} [Y_{T+s} - \mu | \hat{Y}_T] = \lim_{s \rightarrow \infty} \Phi^s \hat{Y}_T = 0$$

- es decir

$$\lim_{s \rightarrow \infty} \mathbb{E} [y_{T+s} | \hat{Y}_T] = \mu$$

en el largo plazo el VAR volverá a su equilibrio estacionario.

Descomposición de la varianza del shock reducido

- Recuerde que los errores reducidos están relacionados con los estructurales por $\epsilon_t = \Gamma_0^{-1} \varepsilon_t$.
- Sea $\Gamma_0^{-1} \equiv A = [a_1 \dots a_n]$, con a_i la i -ésima columna de A .
- Entonces

$$\begin{aligned} \epsilon_t &= a_1 \varepsilon_{1t} + a_2 \varepsilon_{2t} + \cdots + a_n \varepsilon_{nt} \\ \Omega &= \mathbb{E}[\epsilon_t \epsilon_t'] = \sigma_1^2 a_1 a_1' + \sigma_2^2 a_2 a_2' + \cdots + \sigma_n^2 a_n a_n' \\ &= \sum_{j=1}^n \sigma_j^2 a_j a_j' \end{aligned}$$

Ejemplo 10-9: Descomposición de Ω

Tomando Ω del ejemplo 10.1

$$\Omega = \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 4.25 & 2.5 \\ -1 & 2.5 & 12.25 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -1 & 0.75 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix} \begin{smallmatrix} \Gamma_0^{-1} \\ \Sigma \\ \Gamma_0'^{-1} \end{smallmatrix}$$

Entonces $a'_1 = [1 \quad 0.5 \quad -1]$, $a'_2 = [0 \quad 1 \quad 0.75]$, $a'_3 = [0 \quad 0 \quad 1]$ y

$$\begin{aligned} \Omega &= 1 \begin{bmatrix} 1 \\ 0.5 \\ -1 \end{bmatrix} [1 \quad 0.5 \quad -1] + 4 \begin{bmatrix} 0 \\ 1 \\ 0.75 \end{bmatrix} [0 \quad 1 \quad 0.75] + 9 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [0 \quad 0 \quad 1] \\ &= \frac{1}{\sigma_1^2} \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 0.25 & -0.5 \\ -1 & -0.5 & 1 \end{bmatrix} + \frac{4}{\sigma_2^2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0.75 \\ 0 & 0.75 & 0.5625 \end{bmatrix} + \frac{9}{\sigma_3^2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Observe que *esta descomposición depende del ordenamiento de las variables.*

Descomposición de la varianza del pronóstico

- Sustituyendo $\Omega = \sum_{j=1}^n \sigma_j^2 a_j a'_j$ en $\text{Var} [\hat{Y}_{T+s} | \hat{Y}_T]$ tenemos

$$\begin{aligned}\text{Var} [\hat{Y}_{T+s} | \hat{Y}_T] &= \\ &\sum_{j=1}^n \sigma_j^2 \left\{ a_j a'_j + \Phi a_j a'_j \Phi' + \cdots + \Phi^{s-1} a_j a'_j \Phi'^{s-1} \right\}\end{aligned}$$

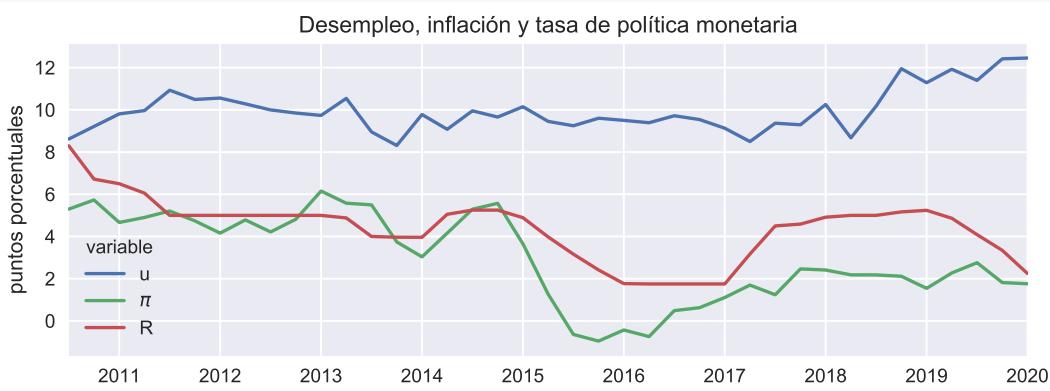
- Con esta expresión, podemos cuantificar la contribución del j -ésimo shock estructural al error cuadrático medio del pronóstico s -períodos adelante.
- Observe que esto asume que conocemos los parámetros del modelo! *No toma en cuenta los errores de estimación.*

Ejemplo 10-10: Un modelo VAR para la política monetaria de Costa Rica

- Supongamos que queremos ver el papel que juega la tasa de política monetaria R del BCCR sobre el desempleo u y la inflación π en nuestro país.
- Para ello estimaremos un VAR:

$$\begin{bmatrix} u_t \\ \pi_t \\ R_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \sum_{i=1}^p \underset{3 \times 3}{\Phi_i} \begin{bmatrix} u_{t-i} \\ \pi_{t-i} \\ R_{t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_t^u \\ \epsilon_t^\pi \\ \epsilon_t^R \end{bmatrix}$$

- Contamos con una muestra de 39 observaciones trimestrales, de 2010-III a 2020-I.



- El primer paso es escoger el número de rezagos p del VAR:

	AIC	BIC	FPE	HQIC
0	1.227	1.360	3.411	1.273
1	-1.779	-1.246*	0.1693	-1.595
2	-2.121*	-1.188	0.1219*	-1.799*
3	-2.007	-0.6741	0.1411	-1.547
4	-2.002	-0.2689	0.1511	-1.404

- En este ejemplo escogeremos un solo rezago, en parte porque tenemos una muestra muy pequeña.

```
Results for equation u
=====
coefficient      std. error      t-stat      prob
-----
const            2.305139       1.265376      1.822       0.069
L1.u             0.708084       0.125437      5.645       0.000
L1.$\pi$          -0.127880      0.087455     -1.462       0.144
L1.R             0.243191       0.124252      1.957       0.050
=====

Results for equation $\pi$
=====
coefficient      std. error      t-stat      prob
-----
const           -0.062525       1.588872     -0.039       0.969
L1.u            -0.003928       0.157505     -0.025       0.980
L1.$\pi$          0.827925       0.109813      7.539       0.000
L1.R             0.119255       0.156017      0.764       0.445
=====

Results for equation R
=====
coefficient      std. error      t-stat      prob
-----
const            2.282890       0.964744      2.366       0.018
L1.u             -0.164187       0.095635     -1.717       0.086
L1.$\pi$          0.091391       0.066677      1.371       0.170
L1.R             0.753419       0.094732      7.953       0.000
=====

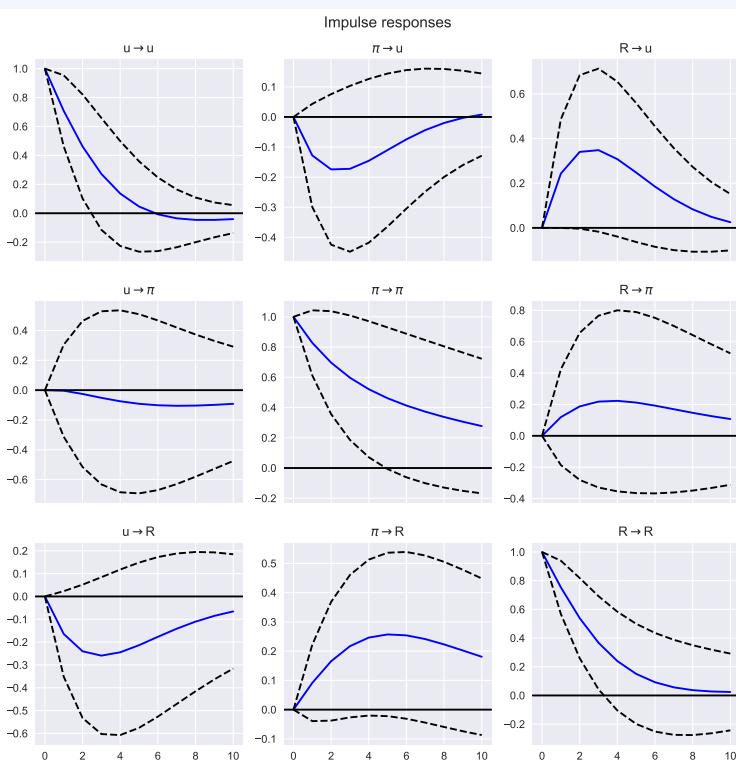
Correlation matrix of residuals
      u      $\pi$         R
u  1.000000 -0.031396 -0.003364
$\pi$ -0.031396  1.000000  0.307269
R   -0.003364  0.307269  1.000000
```

- Al parecer, ninguna variable del sistema causa a otra en el sentido de Granger.

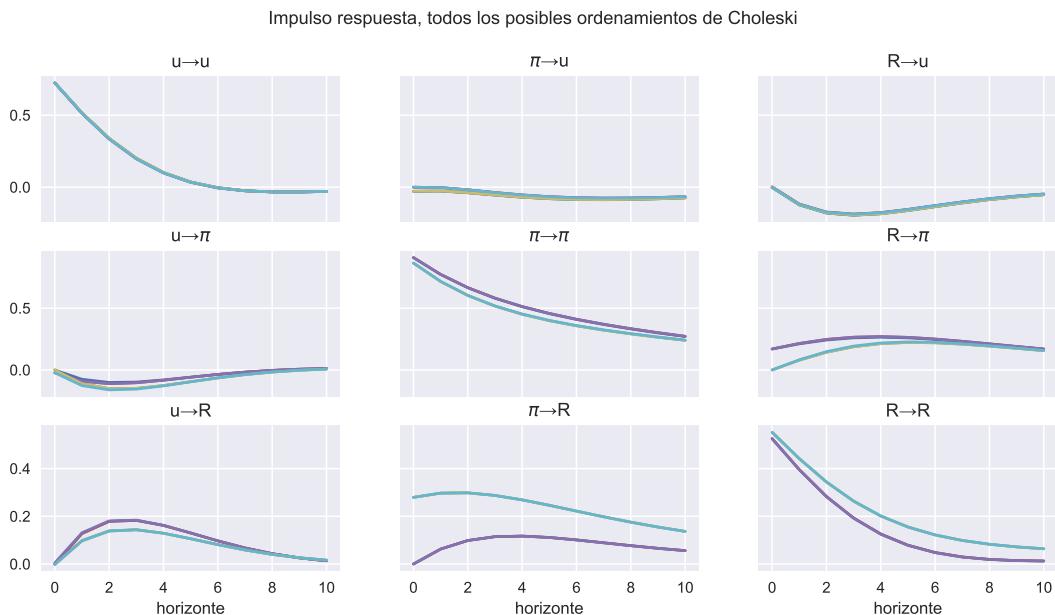
	Dependiente	u	π	R
Explicativa				
u	0.000	0.980	0.089	
π	0.147	0.000	0.173	
R	0.053	0.446	0.000	

Valores p de hipótesis de causalidad.

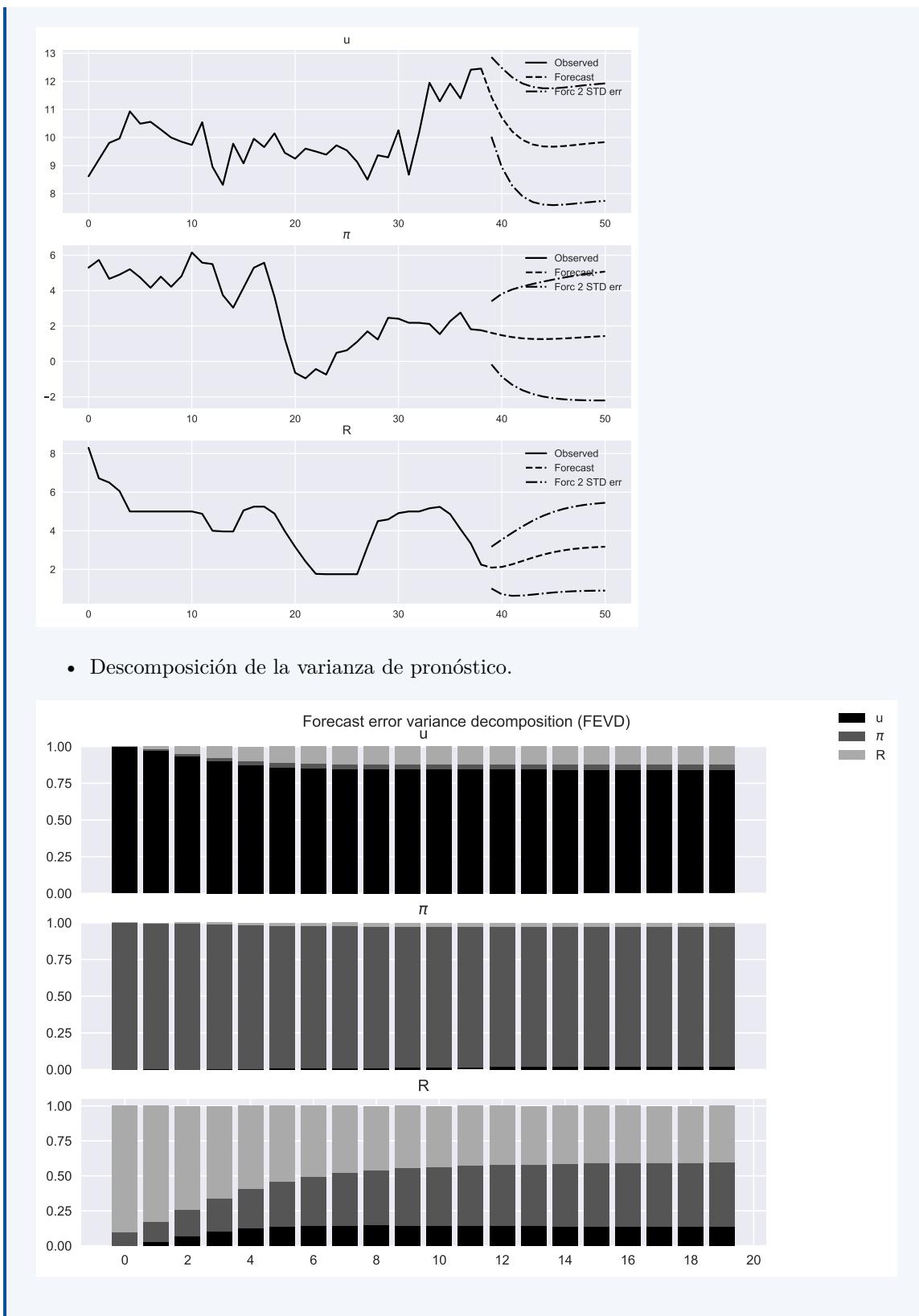
- Esto no necesariamente implica que no haya relación de causalidad entre las variables: podría haber causalidad contemporánea.
- Funciones de impulso respuesta, con impulsos unitarios.



- Funciones de impulso respuesta, con impulsos ortogonales.
- El resultado depende del ordenamiento de las variables en el sistema (Choleski).



- En un VAR estacionario, los pronósticos siempre convergen a la media de largo plazo de cada variable.



Referencias del capítulo 10

- Enders, Walter (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
- Sims, Christopher A. (ene. de 1980). “Macroeconomics and Reality”. En: *Econometrica* 48.1.
- Stock, James H. y Mark W. Watson (2001). “Vector Autoregressions”. En: *Journal of Economic Perspectives* 15.4 (Fall), págs. 101-115.

Capítulo 11

Cointegración y vector de corrección de errores (VECM)

En este capítulo

11.1 Cointegración	187
11.2 Vector de Corrección de Errores (VECM)	191

11.1 Cointegración

Motivation

- A continuación estudiamos la manera de estimar un VAR cuyas series no son estacionarias.
- En procesos univariados, basta con diferenciar la serie y aplicar técnicas de Box-Jenkins.
- En proceso multivariados, es necesario determinar si las series están cointegradas.

Cointegración como equilibrio de largo plazo

- Considere el modelo monetario

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + \epsilon_t$$

oferta demanda brecha

- Para que la noción de equilibrio tenga sentido, la brecha ϵ_t debe ser estacionaria, es decir, $I(0)$
- Esto a pesar de que m_t , p_t , y_t y r_t sean $I(1)$.
- Esto implica que la combinación lineal de variables $I(1)$

$$\begin{bmatrix} 1 & -\beta_1 & -\beta_2 & -\beta_3 \end{bmatrix} \begin{bmatrix} m_t \\ p_t \\ y_t \\ r_t \end{bmatrix} = \beta_0 + \epsilon_t$$

resulta en un proceso $I(0)$

- Decimos que m_t , p_t , y_t y r_t están cointegradas, con vector de cointegración $[1 \quad -\beta_1 \quad -\beta_2 \quad -\beta_3]$.

Cointegración como equilibrio de largo plazo**(cont'n)**

- Teorías de equilibrio con variables no estacionarias requieren la existencia de una combinación lineal de las variables que sea estacionaria
- Otros ejemplos:

- Teoría de la función de consumo:

$$c_t = \beta y_t^p + c_t^t \Rightarrow [1 \quad -\beta] \begin{bmatrix} c_t \\ y_t^p \end{bmatrix} = c_t^t \text{ es estacionario}$$

- Teoría de la paridad del poder de compra:

$$e_t + p_t^* - p_t = [1 \quad 1 \quad -1] \begin{bmatrix} e_t \\ p_t^* \\ p_t \end{bmatrix} \text{ es estacionario}$$

Regresión espuria

- Considere dos caminatas aleatorias independientes

$$\begin{aligned} y_t &= y_{t-1} + u_t & u_t &\text{ ruido blanco} \\ x_t &= x_{t-1} + v_t & v_t &\text{ ruido blanco} \end{aligned}$$

- Como y_t es independiente de x_t , uno esperaría que en la regresión

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

el R^2 y el β_1 tenderían a cero.

- Pero este no es el caso. Con series no estacionarias, la correlación espuria puede persistir aún en muestras grandes.

Ejemplo 11-1: Regresión espuria

Jupyter 05 Relacion espuria.ipynb

Regresión con series I(1):

$$\log(\text{GDP.COSTA.RICA}_t) = \beta_0 + \beta_1 \log(\text{GDP.MALTA}_t) + \epsilon_t$$

OLS Regression Results

Dep. Variable:	CRI	R-squared:	0.806
Model:	OLS	Adj. R-squared:	0.802
Method:	Least Squares	F-statistic:	195.0
Date:	Thu, 02 Jan 2020	Prob (F-statistic):	2.40e-18
Time:	14:07:58	Log-Likelihood:	34.197
No. Observations:	49	AIC:	-64.39
Df Residuals:	47	BIC:	-60.61
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.4814	0.300	14.916	0.000	3.877	5.086
MLT	0.4432	0.032	13.966	0.000	0.379	0.507
<hr/>						
Omnibus:		57.494	Durbin-Watson:		0.064	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		5.218	
Skew:		0.156	Prob(JB):		0.0736	
Kurtosis:		1.432	Cond. No.		164.	
<hr/>						
Regresión con series I(0):						
$\Delta \log(\text{GDP.COSTA.RICA}_t) = \beta_0 + \beta_1 \Delta \log(\text{GDP.MALTA}_t) + \epsilon_t$						
OLS Regression Results						
<hr/>						
Dep. Variable:	CRI	R-squared:			0.032	
Model:	OLS	Adj. R-squared:			0.011	
Method:	Least Squares	F-statistic:			1.537	
Date:	Thu, 02 Jan 2020	Prob (F-statistic):			0.221	
Time:	14:09:30	Log-Likelihood:			102.51	
No. Observations:	48	AIC:			-201.0	
Df Residuals:	46	BIC:			-197.3	
Df Model:	1					
Covariance Type:	nonrobust					
<hr/>						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0143	0.006	2.269	0.028	0.002	0.027
MLT	0.1378	0.111	1.240	0.221	-0.086	0.362
<hr/>						
Omnibus:		27.205	Durbin-Watson:		1.244	
Prob(Omnibus):		0.000	Jarque-Bera (JB):		58.241	
Skew:		-1.586	Prob(JB):		2.26e-13	
Kurtosis:		7.365	Cond. No.		26.4	
<hr/>						

Definición de cointegración (Engle y Granger 1987)

Se dice que los componentes del vector $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ están *cointegrados* de orden (d, b) , denotado por $x_t \sim CI(d, b)$, si

1. Todos los componentes de x_t son integrados de orden d .
2. Existe al menos un vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ tal que la combinación lineal $\beta x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt}$ es integrada de orden $(d - b)$, donde $b > 0$.

A β se le llama *vector de cointegración*

Algunas observaciones acerca de la definición de cointegración

1. Cointegración se refiere a combinaciones *lineales* de variables no estacionarias
2. Si existe, el vector de cointegración no es único
3. Cointegración se refiere a variables del mismo orden; aunque es posible encontrar relaciones de equilibrio entre variables de distinto orden
4. Pueden existir varios vectores de cointegración independientes para un conjunto de variables x_t
5. En la mayor parte de la literatura se entiende cointegración como el caso $CI(1, 1)$.

Pruebas de cointegración: Engle-Granger

Una receta para determinar si las series están cointegradas:

Ingredientes: series de tiempo, software estadístico

Paso 1: Determinar orden de integración de las series

Paso 2: Estimar la relación de equilibrio de largo plazo

Paso 3: Estimar el modelo de corrección de errores

Paso 4: Evaluar si el modelo es adecuado

Pruebas de cointegración de Engle-Granger

En la prueba (aumentada) de Engle y Granger,

$$\Delta \hat{\epsilon}_t = \gamma \hat{\epsilon}_{t-1} + \sum_{i=1}^n a_i \Delta \hat{\epsilon}_{t-i} + \varepsilon_t \quad (\text{nc})$$

$$\Delta \hat{\epsilon}_t = \alpha_0 + \gamma \hat{\epsilon}_{t-1} + \sum_{i=1}^n a_i \Delta \hat{\epsilon}_{t-i} + \varepsilon_t \quad (\text{c})$$

$$\Delta \hat{\epsilon}_t = \alpha_0 + \alpha_1 t + \gamma \hat{\epsilon}_{t-1} + \sum_{i=1}^n a_i \Delta \hat{\epsilon}_{t-i} + \varepsilon_t \quad (\text{ct})$$

$$\Delta \hat{\epsilon}_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \gamma \hat{\epsilon}_{t-1} + \sum_{i=1}^n a_i \Delta \hat{\epsilon}_{t-i} + \varepsilon_t \quad (\text{ctt})$$

si $\gamma = 0$ los residuos $\hat{\epsilon}_t$ presentan raíz unitaria, y por ello las series no estarían cointegradas.

Los valores críticos de MacKinnon

- Para probar la hipótesis nula $H_0 : \gamma = 0$ contra la alternativa $H_1 : \gamma < 0$, se utiliza el estadístico t_γ .
- No obstante, t_γ no tiene la distribución *t-student*, ni siquiera asintóticamente.
- Dado que no es posible derivar la distribución de t_γ analíticamente, es necesario aproximarla con simulaciones de Monte Carlo.
- A partir de tales simulaciones, MacKinnon (2010) presenta valores críticos, que dependen de
 - la especificación determinística (nc, c, ct, ctt),
 - del número de series en el vector de cointegración
 - y del tamaño de muestra T .

Los valores críticos de MacKinnon**(cont'n)**

- Los valores se obtienen evaluando un polinomio en $\frac{1}{T}$

$$C(p) = \beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2} + \beta_3 T^{-3}$$

- Por ejemplo, para probar la cointegración de $N = 3$ variables, con constante y tendencia (ct), la tabla es

Nivel	β_∞	β_1	β_2	β_3
1 %	-4.663	-18.769	-49.793	104.244
5 %	-4.119	-11.892	-19.031	77.332
10%	-3.835	-9.072	-8.504	35.403

- Así, si tenemos 50 observaciones:

$$C(5\%) = -4.119 - \frac{11.892}{50} - \frac{19.031}{50^2} + \frac{77.332}{50^3} = -4.3637$$

Ejemplo 11-2: Valores críticos de Mackinnon

- En el cuaderno de Jupyter Mackinnon [valores críticos para test de cointegración](#) se presentan más ejemplos.
- Se muestra cómo los valores críticos cambian con el tamaño de muestra, el número de series que conforman el vector, y la especificación de los componentes determinísticas de las series.

Nota: Factorización de rango

Suponga que A es una matriz $n \times n$ con rango $r < n$. Existen las matrices X y Y de dimensión $r \times n$ tal que:

$$A = X'Y$$

11.2 Vector de Corrección de Errores (VECM)

De VAR a VECM

$$\begin{aligned}
 y_t &= \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \epsilon_t \\
 \Delta y_t &= (\Phi_1 - I) y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \epsilon_t \\
 &= (\Phi_1 - I) y_{t-1} + (\Phi_2 + \Phi_3) y_{t-2} + \Phi_3 (y_{t-3} - y_{t-2}) + \epsilon_t \\
 &= (\Phi_1 + \Phi_2 + \Phi_3 - I) y_{t-1} + (\Phi_2 + \Phi_3) (y_{t-2} - y_{t-1}) + \\
 &\quad \cdots + \Phi_3 (y_{t-3} - y_{t-2}) + \epsilon_t \\
 &= (\Phi_1 + \Phi_2 + \Phi_3 - I) y_{t-1} - (\Phi_2 + \Phi_3) \Delta y_{t-1} - \Phi_3 \Delta y_{t-2} + \epsilon_t \\
 &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \epsilon_t
 \end{aligned}$$

Si hay $0 < r < n$ vectores de cointegración, entonces Π puede ser descompuesta como el producto de los vectores de cointegración β y los coeficientes de corrección de errores α :

VEC

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \epsilon_t$$

Ejemplo 11-3: Inflación y depreciación en un modelo VEC

- Suponga que en el largo plazo se cumple la PPP: $p_t = e_t + p_t^* + \text{error}_t$ y que las tres variables son I(1) y relacionadas como un VAR(2)
- La representación VECM es

$$\begin{aligned}\Delta p_t^* &= \alpha_1 (p_t - e_t - p_t^*) + \gamma_{11} \Delta p_{t-1}^* + \gamma_{12} \Delta e_{t-1} + \gamma_{13} \Delta p_{t-1} + \epsilon_{1t} \\ \Delta e_t &= \alpha_2 (p_t - e_t - p_t^*) + \gamma_{21} \Delta p_{t-1}^* + \gamma_{22} \Delta e_{t-1} + \gamma_{23} \Delta p_{t-1} + \epsilon_{2t} \\ \Delta p_t &= \alpha_3 (p_t - e_t - p_t^*) + \gamma_{31} \Delta p_{t-1}^* + \gamma_{32} \Delta e_{t-1} + \gamma_{33} \Delta p_{t-1} + \epsilon_{3t}\end{aligned}$$

- o bien

$$\begin{bmatrix} \Delta p_t^* \\ \Delta e_t \\ \Delta p_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \begin{bmatrix} -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_{t-1}^* \\ e_{t-1} \\ p_{t-1} \end{bmatrix} + \Gamma_1 \begin{bmatrix} \Delta p_{t-1}^* \\ \Delta e_{t-1} \\ \Delta p_{t-1} \end{bmatrix} + \epsilon_t$$

- Este modelo explica la inflación internacional, la depreciación, y la inflación doméstica en función de sus propios rezagos y la desviación de los precios y tipo de cambio respecto a su equilibrio de largo plazo.

Pruebas de cointegración: Johansen

- La prueba de Johansen puede verse como una generalización multivariada de la prueba aumentada de Dickey-Fuller
- La prueba y estrategia de estimación permiten estimar *todos* los vectores de cointegración
- Similar a la prueba ADF, la existencia de raíces unitarias implican que la teoría asintótica estándar no es apropiada.

Raíces unitarias: modelo univariado versus multivariado

- Comparemos la prueba ADF con el VECM

$$\begin{aligned}\Delta y_t &= \pi y_{t-1} + \gamma_1 \Delta y_{t-1} + \dots + \gamma_p \Delta y_{t-p} + \epsilon_t && (\text{univariado}) \\ \Delta y_t &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_p \Delta y_{t-p} + \epsilon_t && (\text{multivariado})\end{aligned}$$

- En la prueba ADF, probamos si y_t tiene raíz unitaria con $H_0 : \pi = 0$
- En el caso multivariado, Johansen determina si las series están cointegradas a partir del *rango* de Π

Rango de una matriz vectores de cointegración

Posibles casos del rango:

0: implica $\Pi = 0$, todas las series son I(1) pero no están cointegradas. VAR en diferencias

$0 < r < N$: hay r vectores de cointegración, y escribimos $\Pi = \alpha\beta'$. VECM

N: *cualquier* combinación lineal es estacionaria, lo que implica que las series originales eran estacionarias.

Usando los eigenvalores para determinar el rango

- El rango de una matriz es igual al número de sus eigenvalores distintos de cero.
- Por ello, las pruebas de Johansen están basadas en los eigenvalores de una matriz Π^* semidefinida positiva, *derivada a partir de* Π .
- Suponga que obtenemos Π^* y ordenamos sus eigenvalores de manera tal que

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$$

Prueba de la traza y del máximo eigenvalor

Prueba de la traza

$$\lambda_{\text{traza}}(r) = -T \sum_{i=r+1}^N \ln(1 - \hat{\lambda}_i)$$

Prueba del máximo eigenvalor

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

- Note que $\lambda_i \geq 0 \Rightarrow \ln(1 - \lambda_i) \leq 0$. Ambos estadísticos son no-negativos
- Valores grandes de los estadísticos apuntan a que los eigenvalores son positivos, implicando la existencia de cointegración.

Ejemplo 11-4: Pruebas de Johansen

- Johansen y Juselius (1990) analizan la cointegración de $[m2_t \ y_t \ i_t^d \ i_t^b]$, con datos trimestrales de Dinamarca para el período 1974:1 a 1987:3 (T=53).

H_0	$\hat{\lambda}_i$	λ_{\max}	λ_{traza}
$r = 0$	0.4332	30.09	49.14
$r = 1$	0.1776	10.36	19.05
$r = 2$	0.1128	6.34	8.69
$r = 3$	0.0434	2.35	2.35

Referencias del capítulo 11

- Enders, Walter (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
 Engle, Robert F. y C.W.J. Granger (mar. de 1987). “Co-integration and Error Correction: Representation, Estimation, and Testing”. En: *Econometrica* 55.2, págs. 251-276.
 MacKinnon, James G. (2010). *Critical values for cointegration tests*. Queen's Economics Department Working Paper 1227. Kingston, Ont.

Bibliografía

- Bardsen, Gunnar y col. (2005). *The Econometrics of Macroeconomic Modelling*.
- Beckett, Sean (2013). *Introduction to Time Series Using Stata*. 1^a ed. Stata Press. ISBN: 978-1-59718-132-7.
- Boumans, Marcel J. (15 de feb. de 2015). *Econometrics: The Keynes-Tinbergen Controversy*. URL: <https://dx.doi.org/10.2139/ssrn.2565355>.
- Box, George E. P. y col. (2016). *Time Series Analysis*. 5^a ed. Wiley. ISBN: 978-1-118-67502-1.
- Cogley, Timothy y James M. Nason (1995). “Effects of the Hodrick-Prescott filter on trend and difference stationary time series Implications for business cycle research”. En: *Journal of Economic Dynamics and Control* 19.1–2, págs. 253-278. ISSN: 0165-1889.
- Enders, Walter (2014). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- (2015). *Applied Econometric Time Series*. 4^a ed. Wiley. ISBN: 978-1-118-80856-6.
- Engle, Robert F. y C.W.J. Granger (mar. de 1987). “Co-integration and Error Correction: Representation, Estimation, and Testing”. En: *Econometrica* 55.2, págs. 251-276.
- Greene, William H. (2012). *Econometric Analysis*. 7^a ed. Prentice Hall. ISBN: 978-0-13-139538-1.
- Hamilton, James D. (2017). “Why You Should Never Use the Hodrick-Prescott Filter”. En: *Review of Economics and Statistics* 100.5.
- Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
- Hodrick, Robert J. y Edward C. Prescott (1997). “Postwar U.S. Business Cycles: An Empirical Investigation”. En: *Journal of Money, Credit and Banking* 29.1, págs. 1-16. ISSN: 00222879, 15384616.
- Kirchgässner, Gebhard, Jürgen Wolters y Uwe Hassler (2013). *Introduction to Modern Time Series Analysis*. 2^a ed. Springer. ISBN: 978-3-642-33435-1.
- Levendis, John D. (2018). *Time Series Econometrics. Learning Through Replication*. Springer. ISBN: 978-3-319-98281-6.
- Lucas, Robert E. (1996). “Nobel Lecture: Monetary Neutrality”. En: *Journal of Political Economy* 104.4, págs. 661-682.
- Lucas, Robert Jr (ene. de 1976). “Econometric policy evaluation: A critique”. En: *Carnegie-Rochester Conference Series on Public Policy* 1.1, págs. 19-46.
- Lütkepohl, Helmut y Markus Kratzig, eds. (2004). *Applied Time Series Econometrics*. Cambridge University Press. ISBN: 978-0-521-83919-8.
- MacKinnon, James G. (2010). *Critical values for cointegration tests*. Queen's Economics Department Working Paper 1227. Kingston, Ont.
- Mills, Terence C. (2019). *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*. Academic Press. ISBN: 978-0-12-813117-6.
- Miranda, Mario J. y Paul L. Fackler (2002). *Applied Computational Economics and Finance*. MIT Press. ISBN: 0-262-13420-9.
- Perron, Pierre (nov. de 1989). “The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis”. En: *Econometrica* 57.6, pág. 13611401.
- Sims, Christopher A. (ene. de 1980). “Macroeconomics and Reality”. En: *Econometrica* 48.1.
- Stock, James H. y Mark W. Watson (2001). “Vector Autoregressions”. En: *Journal of Economic Perspectives* 15.4 (Fall), págs. 101-115.

- Valadkhani, Abbas (2004). "History of macroeconometric modelling: lessons from past experience". En: *Journal of Policy Modeling* 26, págs. 265-281.
- Walsh, Carl E. (2010). *Monetary Theory and Policy*. 3^a ed. MIT Press. ISBN: 978-0262-013772.
- Zivot, Eric y Donald W. K. Andrews (jul. de 1992). "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis". En: *Journal of Business & Economic Statistics* 10.3, págs. 251-270.