

Tema 4: Metodología de Box Jenkins (ARIMA)

Rodrigo Ortiz, PhD

Contenidos clase de hoy

1. Introducción
2. LA METODOLOGÍA BOX-JENKINS
3. Modelos autorregresivos
4. Ejemplo Modelos autorregresivos
5. Modelos de promedio móvil
6. Ejemplo Modelos de promedio móvil
7. Método Box-Jenkins

1. Introducción

LA METODOLOGÍA BOX-JENKINS (ARIMA)

- Los modelos de promedio móvil autorregresivo integrado (ARIMA, por sus siglas en inglés) son una clase de modelos lineales que tienen la capacidad de operar sobre series de tiempo **estacionarias o no estacionarias**.
 - Recuerde que los procesos estacionarios varían en torno a un nivel fijo en tanto que los procesos no estacionarios no tienen un nivel promedio constante natural.

1. Introducción

LA METODOLOGÍA BOX-JENKINS (ARIMA)

- Los modelos ARIMA no involucran a las variables independientes en su construcción. En cambio, emplean la información que se encuentra en la serie misma para generar los pronósticos.
 - Por ejemplo, un modelo ARIMA para las ventas mensuales proyectaría un patrón histórico de ventas para producir un pronóstico para las ventas del mes siguiente.
- Los modelos ARIMA dependen mucho de los **patrones de autocorrelación** que existen en los datos. Los estadísticos G. E. P. Box y G. M. Jenkins lograron grandes avances en la metodología para identificar, ajustar y verificar los modelos ARIMA adecuados.
 - Por esto, los modelos ARIMA para producir pronósticos suelen ser llamados metodología Box-Jenkins.

2. LA METODOLOGÍA BOX-JENKINS

- La metodología Box-Jenkins para generar pronósticos es distinta de la mayoría de los métodos debido a que **no supone un patrón** particular en los datos históricos de las series que han de pronosticarse.
- Usa un **método iterativo** para identificar un modelo posible de una clase general de modelos.
- Enseguida, el modelo **seleccionado se contrasta** con los datos históricos para ver si describe con precisión la serie.

2. LA METODOLOGÍA BOX-JENKINS

- El modelo se ajusta correctamente si
 - los residuales son pequeños
 - están distribuidos aleatoriamente
 - no contienen información útil
- Si el modelo especificado **no es satisfactorio**, el proceso **se repite** mediante un nuevo modelo diseñado para mejorar el original.
- Se sigue aplicando este procedimiento iterativo hasta que se encuentra un procedimiento satisfactorio.

2. LA METODOLOGÍA BOX-JENKINS

- La selección inicial de un modelo ARIMA se basa en el examen de una **gráfica de la serie de tiempo** y un examen de su **autocorrelación** para diversos retrasos.
- El patrón de la autocorrelación de la muestra, calculado a partir de la serie de tiempo, coincide con el patrón de autocorrelación ya conocido que se asocia a un modelo ARIMA específico.
 - Este acoplamiento se realiza para las autocorrelaciones y las autocorrelaciones parciales.
- Al seleccionar un modelo, recuerde que las autocorrelaciones calculadas a partir de los datos **no serán exactamente iguales** a las autocorrelaciones teóricas asociadas con un modelo ARIMA.

3. Modelos autorregresivos

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (9.1)$$

donde

Y_t = variable de la respuesta (dependiente) en el tiempo t

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ = variable de respuesta en los retrasos $t-1, t-2, \dots, t-p$,
respectivamente, estas Y desempeñan la función de variables independientes

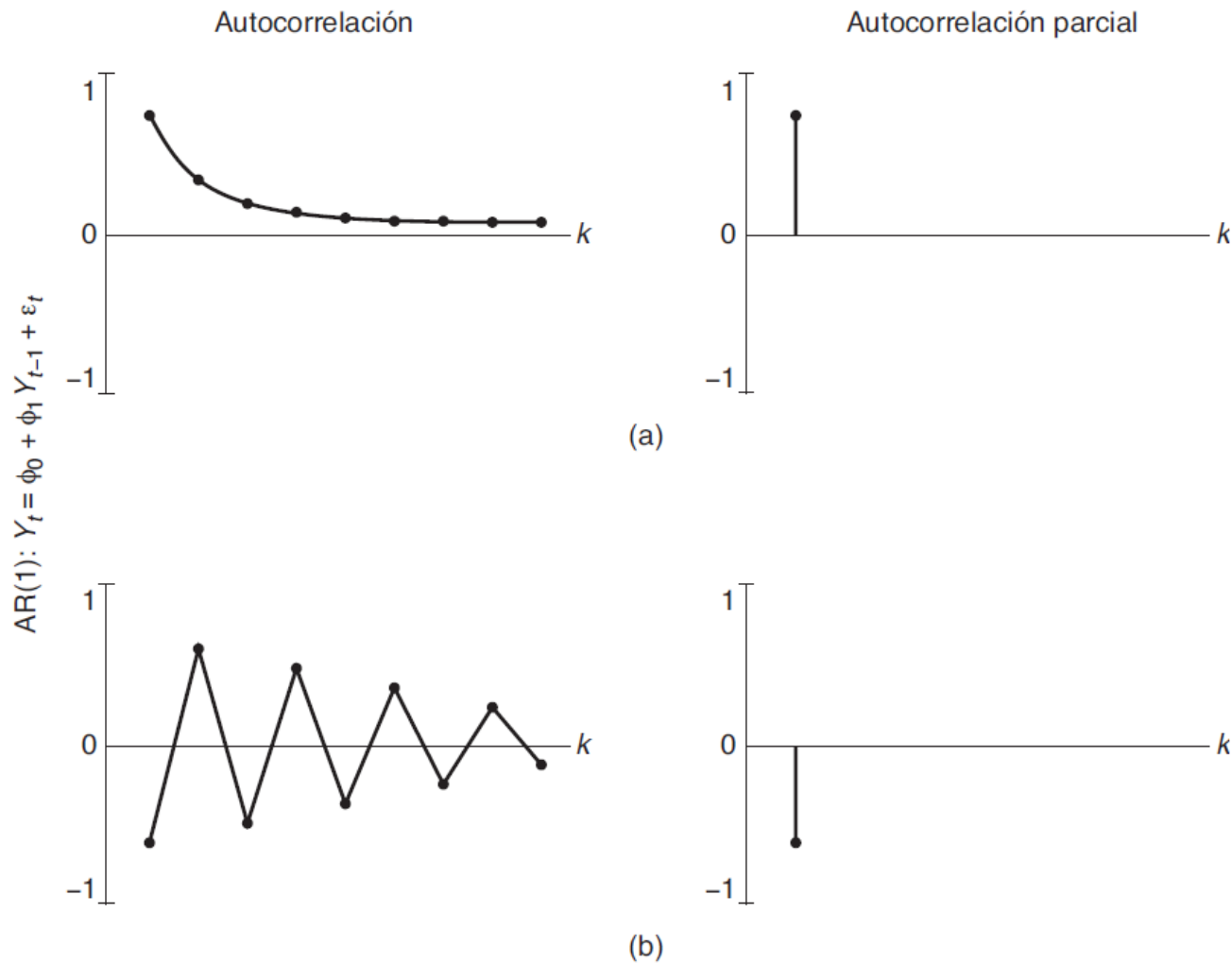
$\phi_0, \phi_1, \phi_2, \dots, \phi_p$ = coeficientes que serán estimados²

ε_t = término de error en el tiempo t que representa los efectos de las variables que no explica el modelo; los supuestos acerca del término de error son las mismas que las del modelo de regresión estándar

3. Modelos autorregresivos

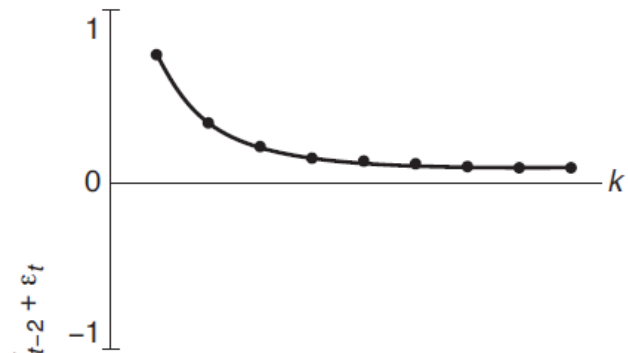
- Los modelos autorregresivos son apropiados para series de tiempo estacionarias y que tienen un coeficiente de ϕ_0 que se relaciona con el nivel constante de la serie.
- Si los datos varían alrededor de cero o se expresan como desviaciones de la media $Y_t - \bar{Y}$, no se requiere el coeficiente ϕ_0 .

3. Modelos autorregresivos

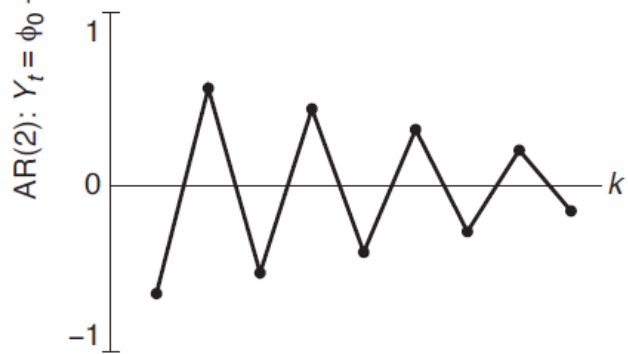
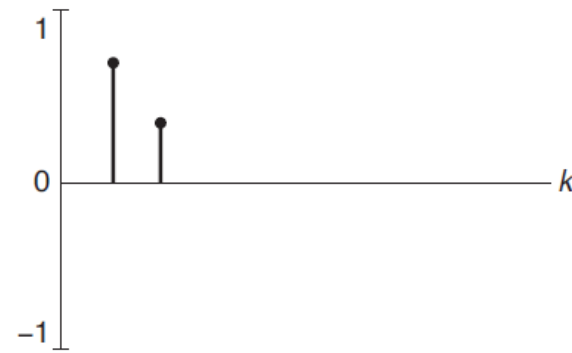


Los coeficientes de autocorrelación se aproximan gradualmente al cero, a la vez que los coeficientes de autocorrelación parcial caen a cero después del primer retraso de tiempo.

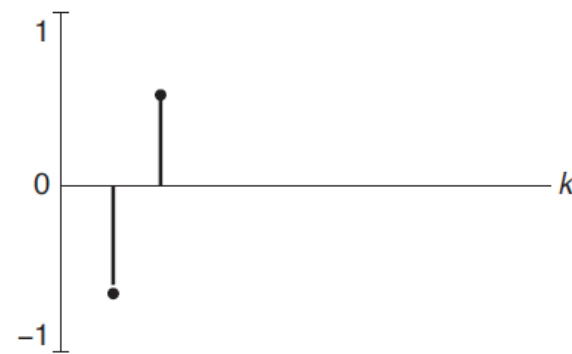
3. Modelos autorregresivos



(c)



(d)



De nuevo, los coeficientes de autocorrelación se aproximan a cero y los coeficientes de autocorrelación parcial caen a cero después del segundo tiempo de retraso.

4. Modelos de promedio móvil

Un modelo de promedio móvil de orden q adopta la forma

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (9.2)$$

donde

Y_t = variable de respuesta (dependiente) en un tiempo t

μ = valor promedio que permanece constante en el proceso

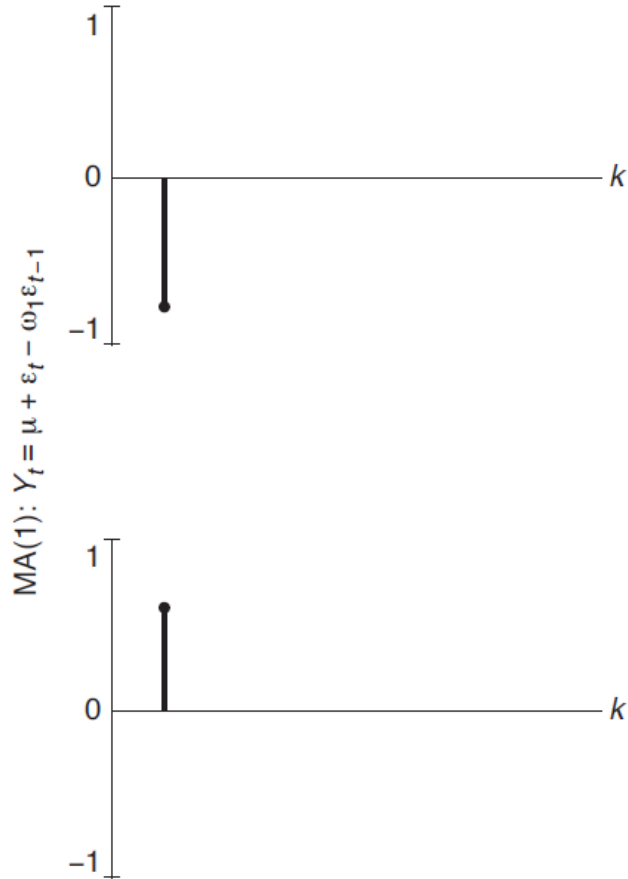
$\omega_1, \omega_2, \dots, \omega_q$ = coeficientes que serán estimados

ε_t = término de error que representa los efectos de las variables no explicadas por el modelo; los supuestos acerca del término de error son las mismas que los del modelo de regresión estándar

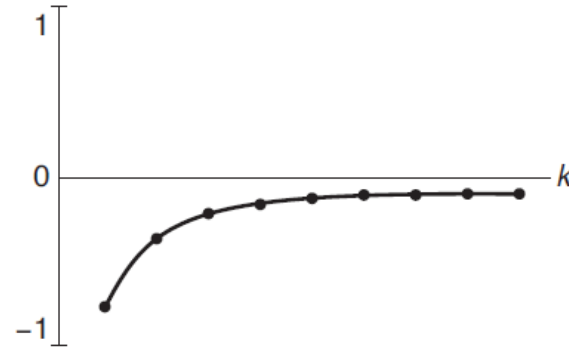
$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = errores en periodos anteriores al tiempo t , incorporados en la respuesta Y_t .

4. Modelos de promedio móvil

Autocorrelación

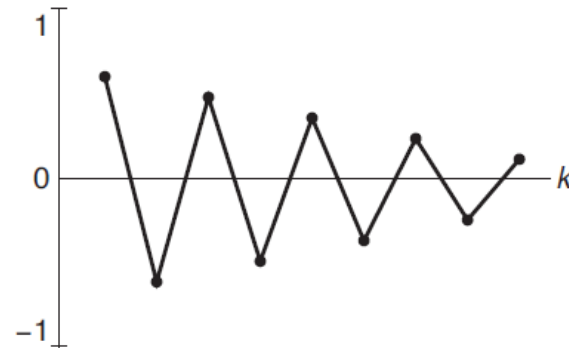


Autocorrelación parcial



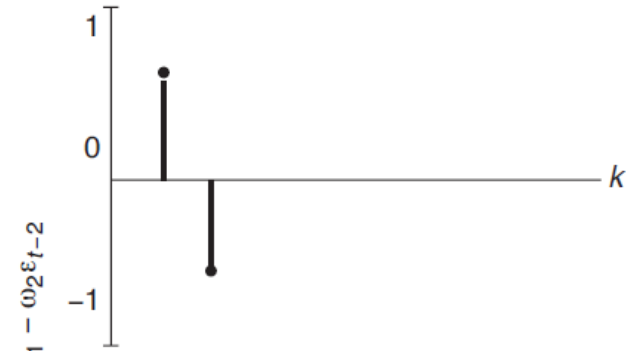
(a)

Los coeficientes de autocorrelación en el modelo MA(1) caen a cero después del primer retraso de tiempo, mientras que los coeficientes de autocorrelación parcial también se aproximan al cero pero lo hacen gradualmente.

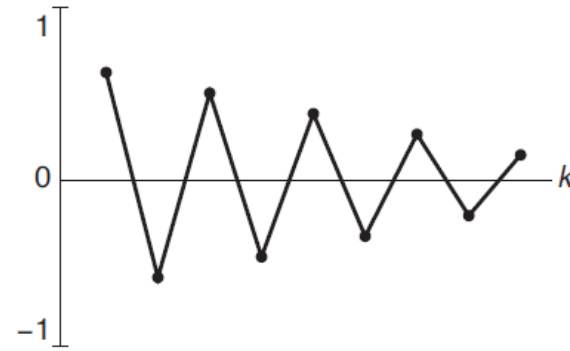


(b)

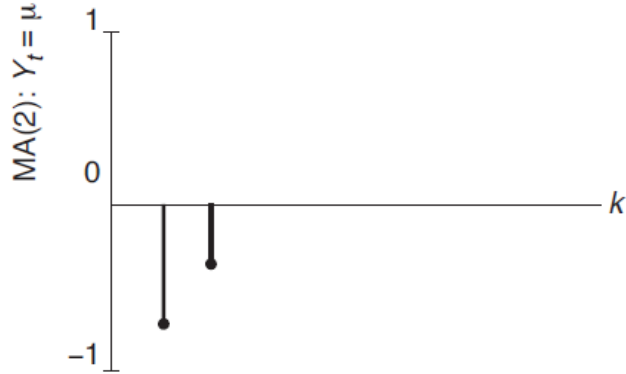
4. Modelos de promedio móvil



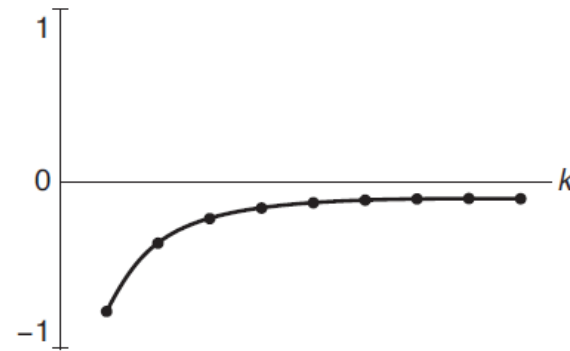
(c)



Además, los coeficientes de autocorrelación del modelo MA(2) equivalen a cero después del segundo retraso de tiempo, mientras que los coeficientes de autocorrelación parcial se aproximan a cero gradualmente.



(d)



6. Método Box-Jenkins

Paso 1: identificación del modelo

- 1. El primer paso en la identificación del modelo es determinar si la serie es estacionaria; es decir, si la serie de tiempo aparenta variar alrededor de un nivel fijo.
 - Una serie de tiempo no estacionaria se indica si la serie parece crecer o decrecer con relación al tiempo y las autocorrelaciones no pueden desvanecerse con rapidez.

Paso 1: identificación del modelo

Ejemplo: Serie no estacionaria

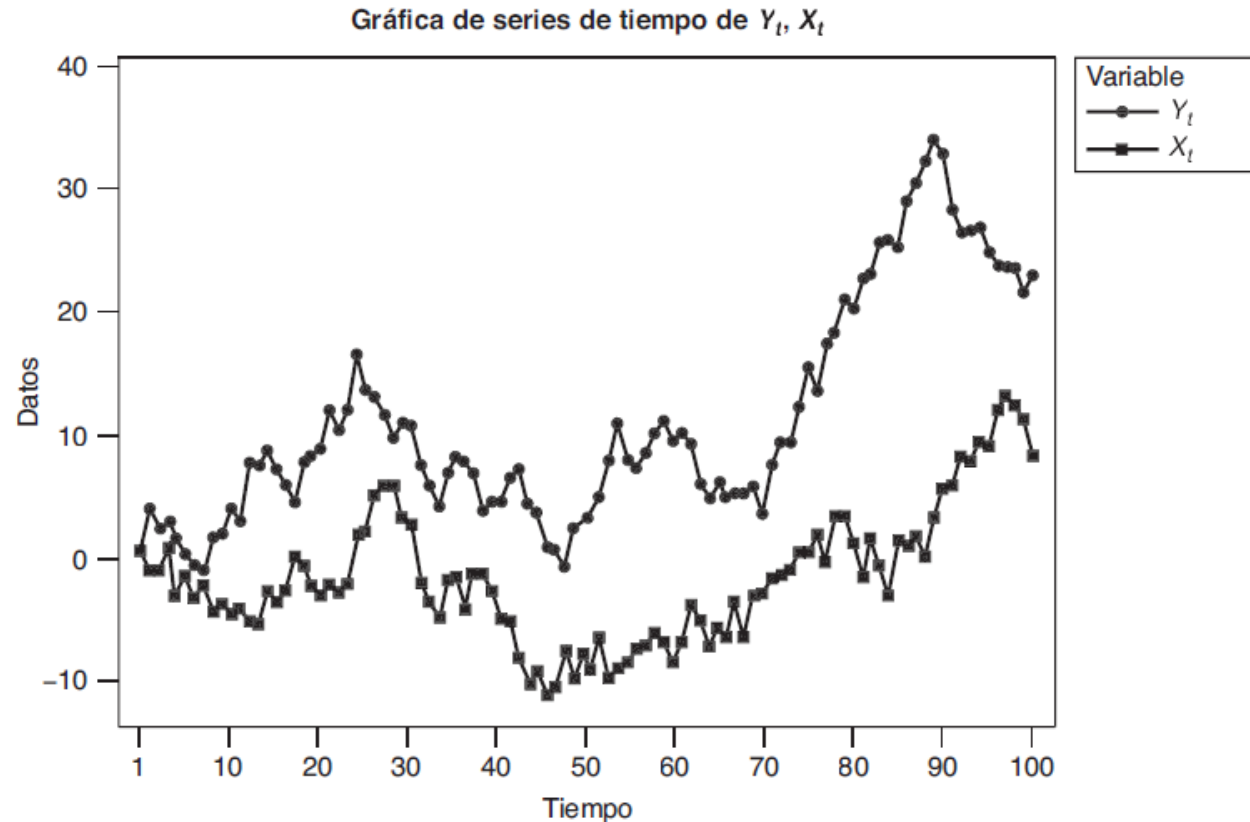


FIGURA 8.2 Gráficas de series de tiempo de dos series no relacionadas, Y_t (superior) y X_t (inferior).

Paso 1: identificación del modelo

Primera diferencia

- Si la serie no es estacionaria, con frecuencia puede convertirse en una serie estacionaria al tomar sus diferencias.

Por ejemplo, suponga que la serie original Y_t por lo general se incrementa con el tiempo, pero las primeras diferencias $\Delta Y_t = Y_t - Y_{t-1}$ varían alrededor de un nivel fijo. Podría ser apropiado modelar las diferencias estacionarias por medio de un modelo ARMA, de, por ejemplo, un orden $p = 1$ y $q = 1$. En este caso el modelo es⁵

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

o

$$(Y_t - Y_{t-1}) = \phi_1 (Y_{t-1} - Y_{t-2}) + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

Paso 1: identificación del modelo

ARIMA

- Se toman diferencias hasta que la gráfica de los datos indica que
 - la serie varía alrededor de un nivel fijo y
 - las autocorrelaciones de la muestra desaparecen con rapidez
- El número de diferencias requerido para lograr un estado estacionario se denota por la **d**.
- A los modelos para las series que no son estacionarias se les llama **modelos de promedio móvil integrados autorregresivos** y se denotan como ARIMA (p, d, q).

Paso 1: identificación del modelo

ARIMA

- En este caso, p indica el orden de la parte autorregresiva, d indica el orden de la diferencia y q , el orden de la parte de promedio móvil.
- Si la serie original es estacionaria, entonces $d = 0$ y los modelos ARIMA se reducen a modelos ARMA.
- En consecuencia, desde este punto, la notación ARIMA (p, d, q) se utiliza para indicar los modelos tanto para series de tiempo estacionarias ($d = 0$) como para las no estacionarias ($d > 0$).

Test para raíz unitaria

- Prueba Dickey-Fuller
 - H_0 : No es estacionaria (raíz unitaria)
 - H_1 : Es estacionaria
- STATA `dfuller`
- R `adf.test`

Paso 1: identificación del modelo ARIMA

- 2. Una vez que se ha obtenido una serie estacionaria, el analista debe **identificar la forma del modelo** que habrá de utilizar.
- La segunda parte del paso 1 se consigue al comparar la autocorrelación y la autocorrelación parcial que se calcularon a partir de los datos para las autocorrelaciones autocorrelaciones parciales teóricas de los diversos modelos ARIMA.

Paso 2: estimación de modelos

- Es probable que haya cierta ambigüedad al determinar un modelo ARIMA apropiado a partir de los patrones que provienen de las autocorrelaciones y las autocorrelaciones parciales de la muestra.
 - De esta manera, la selección del modelo inicial deberá considerarse como tentativa. Los análisis pueden realizarse en los pasos 2 y 3 para determinar si el modelo es adecuado. Si no fuera el caso, se deberá intentar con un modelo alternativo.
 - Con un poco de práctica, el analista deberá volverse un experto para identificar un modelo adecuado.
- Una vez que se ha seleccionado un modelo tentativo, deben **estimarse** los parámetros para dicho modelo.

Paso 2: estimación de modelos

Además, se calcula el *error cuadrado medio de los residuales*, un estimado de la varianza del error ε_t .

El *error cuadrado medio de los residuales* se define como⁷

$$s^2 = \frac{\sum_{t=1}^n e_t^2}{n - r} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n - r} \quad (9.4)$$

donde

$e_t = Y_t - \hat{Y}_t$ = el residual en el tiempo t

n = el número de residuales

r = el número total de los parámetros estimados

Paso 3: evaluación del modelo

1. Muchas de las gráficas de los residuales que son útiles para el análisis de regresión pueden desarrollarse para los residuales de un modelo ARIMA.
2. Las autocorrelaciones residuales individuales deberán ser pequeñas y, por lo general, estar dentro de $\pm 2/\sqrt{n}$ de cero. Las autocorrelaciones residuales significativas en retrasos cortos o estacionales sugieren que el modelo no es adecuado y que se debe elegir un modelo nuevo o modificado.
3. Como un grupo, las autocorrelaciones residuales deberán ser coherentes con aquellas producidas por los errores aleatorios.

Paso 3: evaluación del modelo

- Una prueba chi cuadrada (χ^2) que se basa en la estadística de Ljung-Box Q proporciona una revisión global de la pertinencia del modelo. Esta prueba considera las dimensiones de las autocorrelaciones residuales como un grupo. La estadística de prueba Q es

$$Q_m = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k} \quad (9.5)$$

la cual se distribuye aproximadamente como una variable aleatoria de chi cuadrada con grados de libertad $m - r$ en donde r es el número total de parámetros estimados en el modelo ARIMA. En la ecuación 9.5,

$r_k(e)$ = la autocorrelación residual en el retraso k

n = el número de residuales

k = el retraso de tiempo

m = el número de retrasos de tiempo que habrán de ser evaluados

Prueba de Ljung-Box

- Esta prueba permite probar en forma conjunta de que todos los coeficientes de autocorrelación son simultáneamente iguales a cero, esto es que son independientes, está definida como

$$LB = n(n + 2) \sum_{k=1}^m \left(\frac{\widehat{\rho}_k^2}{n - k} \right) \sim \chi^2_{(m)}$$

- Donde n tamaño de la muestra, m longitud del rezago
 - H_0 : Las autocorrelaciones son independientes
 - H_1 : Las autocorrelaciones no son independientes
- En una aplicación, si Q calculada excede el valor de Q *crítico* de la tabla *ji cuadrada* al nivel de significancia seleccionado, no se acepta la hipótesis nula de que todos los coeficientes de autocorrelación son iguales a cero; por lo menos uno de ellos deben ser diferentes de cero.

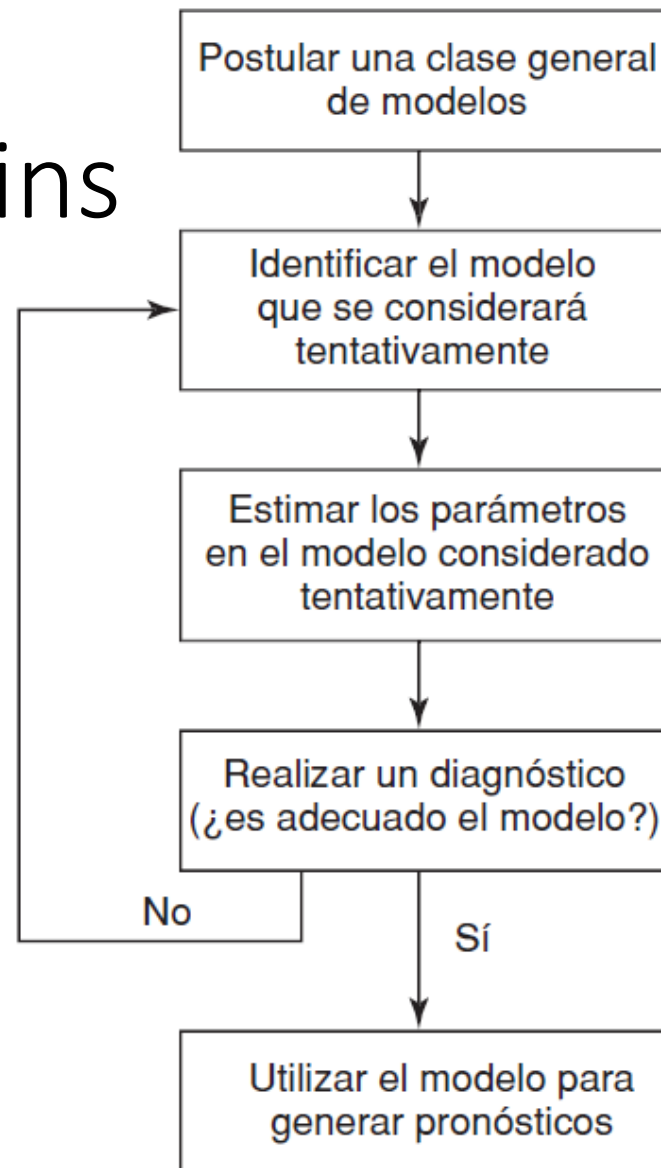
Paso 3: evaluación del modelo

- Si el valor p asociado con la estadística Q es pequeño (por ejemplo, un valor p de $<.05$), se considera que el **modelo es inadecuado**.
 - El analista deberá considerar un modelo nuevo o modificado y continuar el análisis hasta que se determine un modelo satisfactorio.

Paso 4: realización de pronósticos con el modelo

- 1. Después de que se ha encontrado un modelo adecuado, se pueden llevar a cabo los pronósticos para un periodo, o varios, en el futuro.
- 2. A medida que se tienen más datos disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos revisados que procedan de otro origen de tiempo.
- 3. Si el patrón de la serie parece cambiar con el tiempo, los nuevos datos podrían usarse para volver a estimar los parámetros del modelo o, de ser necesario, desarrollar un modelo completamente nuevo.

7. Método Box-Jenkins



Fuente: Box, G. E. P., Jenkins, G. M. y G. C. Reinsel, *Time Series Analysis, Forecasting and Control* tercera edición, Upper Saddle River, Nueva Jersey, Prentice Hall, 1994, p. 17. Reimpreso con autorización.

FIGURA 9.1 Diagrama de flujo de la estrategia de construcción del modelo Box-Jenkins

8. Ejemplos ARIMA - Box-Jenkins

- A trabajar...