Tema 4: Modelos Univariantes de Series de Tiempo Estacionarias

Serafín Frache

2016

Datos de series temporales

- En economía gran parte de los datos utilizados son series de tiempo.
    - PBI de Uruguay 1983-2011
    - Tasa de desempleo de Uruguay 1983-2011
    - Precio del Petróleo mensual 2000-2012
    - Ventas diarias de un comercio en el primer semestre 2011
    - Promedio por minuto del índice Dow Jones el 22 de mayo de 2012
    - Número de pasajeros mensual de una aerolínea

Datos de series temporales

- No se puede interpretar que los datos son N observaciones independientes de una misma variable aleatoria, como es el caso con datos de Corte Transversal:
  - Al cambiar el momento de observación cada dato proviene de una variable aleatoria específica para el período.
  - Además, el dato de un período en general no es independiente de los valores observados en períodos anteriores.

Definición

Serie de Tiempo (1): Conjunto de observaciones de una variable X tomadas a intervalos regulares de tiempo.

Componentes de una Serie de Tiempo

- Y_t puede tener un componente determinista y otro aleatorio (''shocks" o ''innovaciones").
  Por ejemplo, si éstos se combinan aditivamente:

$$Y_t = D_t + A_t$$

  - '*Hora a la que sale el sol*' es una ST puramente determinista
  - '*Primer número del sorteo del 5 de Oro*' es una ST puramente aleatoria

- Las series económicas suelen presentar ambos componentes, y en particular se caracterizan por que sus valores están 'serialmente correlacionados'

Enfoques para la estimación

- ¿Qué variables se asume que componen D_t?
  - Exclusivamente valores retardados de $Y \implies$ MODELOS UNIVARIANTES (*Tema 4*)
  - Exclusivamente valores contemporáneos de $X_1, X_2, ..., X_k \implies$ MODELOS ESTÁTICOS (*Tema 5*)
  - Valores contemporáneos y retardos de $X_1, X_2, ..., X_k \implies$ MODELOS DE RETARDOS DISTRIBUIDOS (*Tema 5*)
  - Valores contemporáneos y retardos de $X_1, X_2, ..., X_k$ y valores retardados de $Y \implies$ MODELOS DINÁMICOS (*Temas 5 y 6*)
- PERO: Los resultados de las regresiones anteriores pueden ser disparatados si no se considera adecuadamente una característica central en los datos de series temporales: la *ESTACIONARIEDAD* o *NO ESTACIONARIEDAD* de las series en consideración.

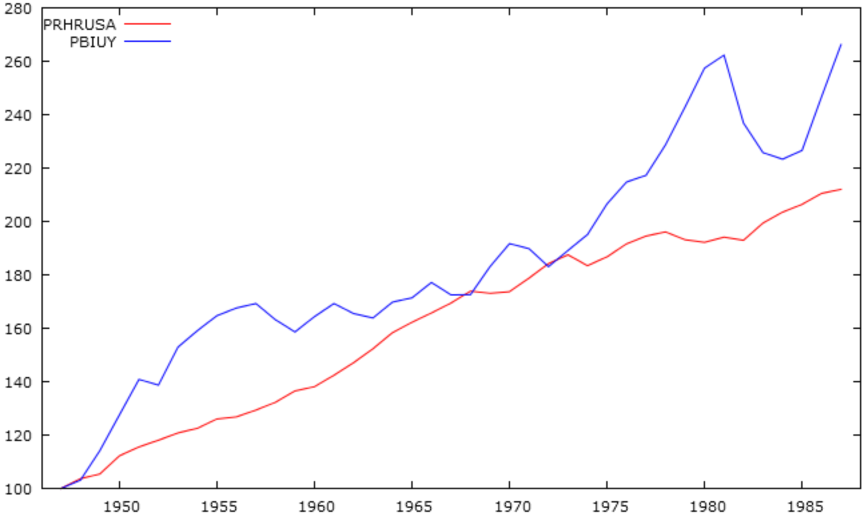
Conceptos básicos: Tendencias y regresión espuria

- Las series económicas suelen describir algún tipo de tendencia
- La regresión simple entre series con tendencia podrá ser 'espuria'
  - La tendencia hará que sus valores contemporáneos estén altamente correlacionados, incluso cuando no haya una relación causal entre las variables.
  - Sin embargo, en sí mismo esto no lleva a que se violen los supuestos clásicos
- Ejemplo de regresión espuria: Modelo 'sin sentido' para el PBI de Uruguay:
  - MRLS sobre el producto por hora de trabajo en EEUU

$$PBIUY_t = \beta_0 + \beta_1 PRHRUSA_t + u_t$$

  - Las tendencias en ambas variables llevarán a $\widehat{\beta}_1$ significativo y R^2 alto

Conceptos básicos: Tendencias y regresión espuria (2)



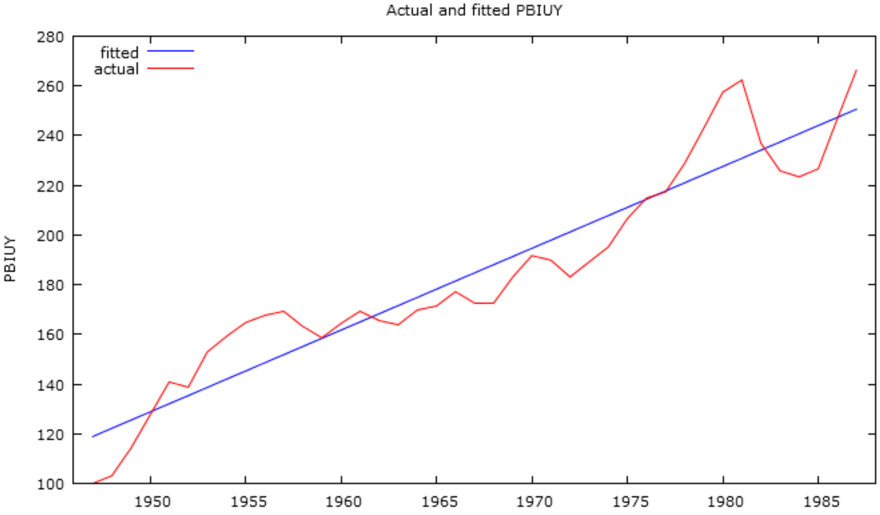
Conceptos básicos: 'Detrending'

- **A**- Una estrategia posible para realizar la estimación es quitar a cada serie su tendencia ('detrend').
  - Implica suponer que el valor medio es una función en t:
  - Tomando una tendencia lineal:

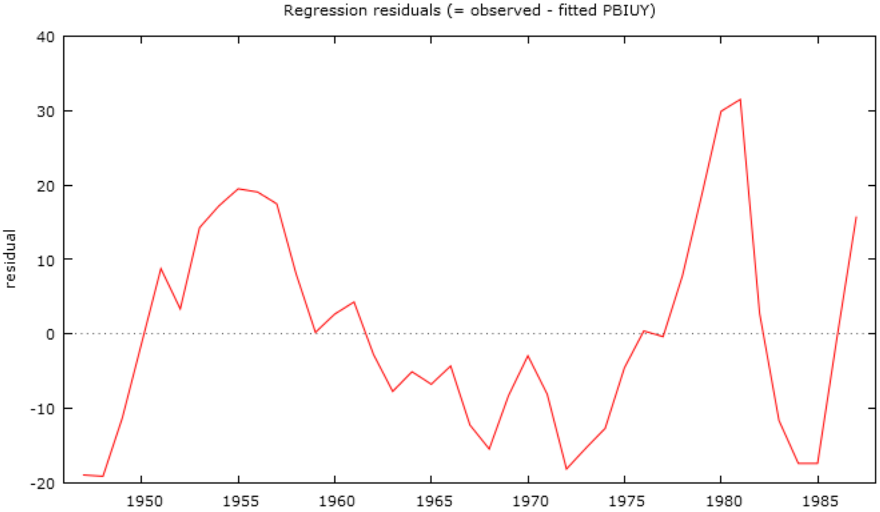
$$E(Y_t) = \alpha_0 + \alpha_1 t$$

  - Al estimar el modelo: $Y_t = \alpha_0 + \alpha_1 t + e_t$
    $\{e_t\}$ puede ser utilizada para la estimación como la serie Y_t a la que se le ha eliminado la tendencia lineal $\alpha_1 t$.
  - Si $\{e_t\}$ es es una sucesión i.i.d., entonces $\{Y_t\}$ es una sucesión independiente pero no idénticamente distribuida (¿por qué?)
  - Notar: En sí misma la presencia de una tendencia en la serie no lleva a la violación de ninguno de los supuestos clásicos.
  - En general $\{e_t\}$ no será i.i.d., presentando 'autocorrelación'.

Conceptos básicos: 'Detrending'



Conceptos básicos: 'Detrending'



Conceptos básicos: Inclusión de tendencia temporal

- **B-** Inclusión de una tendencia en el modelo original.

$$PBIUY_t = \beta_0 + \beta_1 PRHRUSA_t + \beta_2 t + u_t$$

  - La variable t recoge factores no observables que crecen o decrecen regularmente con el tiempo.
  - Aquí vemos que la omisión de la variable t ocasiona los problemas de sesgo propios de la omisión de variables relevantes.
  - Esos problemas son más graves en el caso en que la variable omitida esté altamente correlacionada con los regresores (es decir, en el caso en que los regresores presentan una tendencia lineal).
  - Los coeficientes estimados en el modelo original con tendencia son exactamente iguales a los obtenidos en una regresión donde a todas las variables se les ha eliminado la tendencia.
  - En este ejemplo, $\widehat{\beta}_1$ será no significativo

Conceptos básicos: Tendencias no lineales

- Muchas series económicas pueden requerir tendencias no lineales
  - Pueden considerarse **tendencias polinómicas**
    Las variaciones respecto a una tendencia cuadrática serán los residuos de la regresión:
    $$Y_t = \alpha_0 + \alpha_1 t + \alpha_3 t^2 + e_t$$
  - El caso más común es el de **tendencia exponencial**, que se da cuando una serie tiende a crecer o decrecer a una tasa media constante
    En este caso, las variaciones respecto a la tendencia serán los residuos de la regresión:
    $$Ln(Y_t) = \alpha_0 + \alpha_1 t + e_t$$

Conceptos básicos: Tendencia exponencial

- Recordando que para variaciones pequeñas:

$$\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1}) \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

  - Haciendo $\Delta e_t = 0$ se tiene que:

$$\alpha_1 \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

  - Por lo que $\hat{\alpha}_1$ brinda una estimación de la tasa de crecimiento media

Nota: La aproximación se obtiene por un desarrollo de Taylor en torno al valor 1 (ver Uriel pp. 62 y 63)

Conceptos básicos: R-cuadrado en modelos con tendencia

- En series temporales el R^2 es generalmente más alto que con datos de corte transversal
  - En parte esto se debe a que en general se trabaja con datos agregados
  - Pero además, si se requiere tendencia, el R^2 será artificialmente alto
- Recordando la fórmula del R^2:

$$R^2 = 1 - \frac{SCR}{SCT}$$

  - la tendencia participa en $\widehat{y}_t$, por lo que los residuos la 'tienen en cuenta'
  - la SCT no la tiene en cuenta, por lo que se sobreestima $V\widehat{ar}(y_t)$ y el R^2 es artificialmente alto
- Para obtener un R^2 que corrija este problema se estima una regresión eliminando la tendencia a la variable dependiente

Conceptos básicos: Transformación en primeras diferencias

- **C**- Una transformación de los datos que -entre otras cosas- elimina la tendencia es el uso de **primeras diferencias**

$$\Delta Y_t = Y_t - Y_{t-1}$$

  - En el caso en que una serie presenta tendencia:

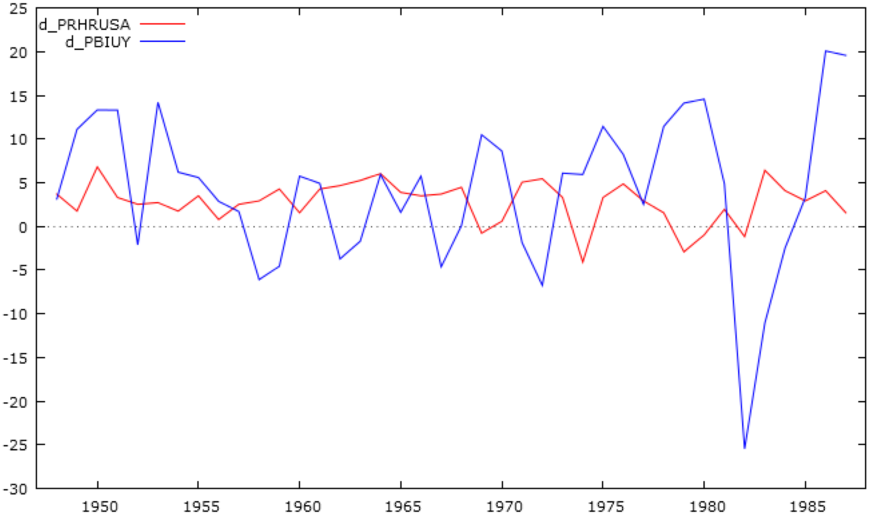
$$Y_t = \alpha_0 + \alpha_1 t + e_t$$

    La primera diferencia será:

$$\Delta Y_t = Y_t - Y_{t-1} = \alpha_1 + e_t - e_{t-1} = \alpha_1 + \Delta e_t$$

  - La serie ΔY_t es la constante α_1 que recoge la variación media entre dos observaciones sucesivas, y Δe_t que son las primeras diferencias entre los valores de la variable Y_t una vez que se eliminó la tendencia
  - $E(\Delta Y_t) = \alpha_1$: Una constante significativa en el modelo en diferencias equivale a una tendencia significativa en el modelo en niveles.

Conceptos básicos: Transformación en primeras diferencias



Conceptos básicos: El operador de retardos

- En series temporales es práctico utilizar el **operador de retardos:**
  - Se trata de un operador lineal tal que:

$$L^k.Y_t \equiv Y_{t-k}$$

  - Algunas propiedades que serán útiles:

$$Lc = c$$

$$L^{-k}.Y_t \equiv Y_{t+k}$$

$$(L^k + L^j)Y_t \equiv L^k Y_t + L^j Y_t = Y_{t-k} + Y_{t-j}$$

- Por lo tanto:

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$$

Nota: Las propiedades del operador de retardos están presentadas con detalle en Enders, Cap. 1, Secc. 9 (todas las ediciones) o Uriel, Apéndice 1.

Conceptos básicos: Especificación dinámica de un modelo

- Un dato económico suele estar correlacionado con sus valores pasados
  - Recordando que Y_t puede verse como:

    $$Y_t = D_t + A_t$$

    Se dirá que el modelo tiene una **especificación dinámica completa** en el caso en que D_t recoja todos los retardos necesarios de la variable dependiente, de los regresores y del shock.
- Cuando la especificación dinámica es completa los residuos no mostrarán autocorrelación serial
- Si existe autocorrelación en los residuos se está violando uno de los supuestos de Gauss-Markov, y por lo tanto los estimadores MCO no son MELI, invalidando la inferencia incluso asintóticamente

Detección de la autocorrelación: t-test para residuos AR(1)

- Partiendo de un modelo como: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$
  - Suponemos que se verifican los restantes supuestos de Gauss-Markov, y en particular que las variables explicativas son estrictamente exógenas ($E(u_t|X_{1,t-1}, X_{2,t-1}, X_{1,t-2}, ...) = 0$)
  - Una forma directa de detectar *asintóticamente* la existencia de autocorrelación de primer orden del término de error, parte de la siguiente regresión:

$$u_t = \rho u_{t-1} + e_t$$

  - Suponiendo que $E(e_t|u_{t-1}, u_{t-2}, ...) = 0$ y que $Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2$, puede probarse $H_0 : \rho = 0$ por medio del estadístico t habitual para $\hat{\rho}$
  - Dado que u_t es inobservable, puede ser remplazada por $\hat{u}_t$
  - Conviene utilizar estimadores robustos a la heteroscedasticidad
  - La utilidad de contrastes asintóticos en Series Temporales es limitada por la disponibilidad de información

Detección de la autocorrelación: Durbin-Watson

- La forma más frecuente de detectar autocorrelación de primer orden en los residuos es el **Contraste de Durbin-Watson**:

$$DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

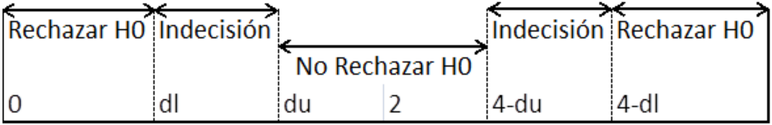
- Puede demostrarse que está relacionado con la prueba anterior, ya que:

$$DW \approx 2(1 - \hat{\rho})$$

  - Durbin y Watson derivaron la distribución del estadístico, y los valores críticos se pueden consultar en sus tablas. Estos dependen del tamaño muestral, del número de regresores y de la existencia de un término constante. Gretl presenta directamente el p-valor de la prueba.

Detección de la autocorrelación: Durbin-Watson

- La lectura de los valores críticos de la distribución DW es menos directa que en otros casos:
  - El estadístico varía entre 0 y 4
  - Valores cercanos a 2 llevarán a No Rechazar H_0
  - Valores cercanos a 4 o a 0 llevarán a Rechazar H_0
  - Sin embargo, existen regiones intermedias en que el valor del estadístico no permite concluir
  - En las tablas se dispone de un valor d_L y un valor d_U
  - Las regiones no concluyentes son: $d_L < DW < d_U$ y $(4 - d_U) < DW < (4 - d_L)$



Variables ficticias estacionales

- Cuando una serie económica tiene una frecuecia mayor que anual (mensual o trimestral, p.e.) suele presentar **estacionalidad**
  - Igual que sucede con las tendencias, hay distintos tipos de estacionalidad
  - La 'estacionalidad determinista' puede modelarse muy facilmente incorporando **variables ficticias estacionales**
  - En una serie mensual, ello implica incorporar binarias para 11 de los meses del año a un modelo con constante
  - Si se hace una regresión de la serie contra 12 binarias mensuales (sin constante) se obtendrá la serie 'desestacionalizada'
  - Las doceavas diferencias también eliminan la estacionalidad: $\Delta^{12} Y_t = Y_t - Y_{t-12} = (1 - L^{12}) Y_t$
  - Al final de este capítulo consideraremos otras posibilidades para el tratamiento de la estacionalidad

Procesos Estocásticos - Definiciones Básicas

Definición

Proceso estocástico: Sucesión de variables aleatorias ordenadas de acuerdo a un índice que varía en (-∞, $+\infty$). Más formalmente: Sea S el conjunto de eventos posibles y T un subconjunto de números reales, se define un proceso estocástico como la secuencia ordenada de variables aleatorias $\{Y_t, t \in T\}$ con $y : S \times T \longrightarrow \mathbb{R}$

Definición

Muestra: Secuencia ordenada de variables aleatorias pertenecientes al intervalo $(1, T)$

Definición

Serie de Tiempo (2): Una realización concreta de dicha muestra (de tamaño 1, salvo que se controlen las condiciones del 'experimento')

Procesos Estocásticos - Caracterización

- Un proceso estocástico (P.E.) está '**perfectamente caracterizado**' cuando se conoce la distribución conjunta: $F(Y_1, Y_2, ..., Y_T)$
- Sin embargo, las características más relevantes del P.E. pueden analizarse a partir sus momentos de primer y segundo orden:

$$\mu_t = E(Y_t)$$

$$\gamma_{t,s} = Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)]$$
$$\gamma_{t,t} = Var(Y_t) = E(Y_t - \mu_t)^2$$

- El coeficiente de correlación (no afectado por unidades de medida) es:

$$\rho_{t,s} = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t) \cdot Var(Y_s)}} \quad donde \quad -1 \leq \rho_{t,s} \leq 1$$

Procesos Estocásticos - Estacionariedad Fuerte

- La estimación de los momentos de un P.E. es simple cuando éste cumple la propiedad de *ESTACIONARIEDAD*
- Se dice que un P.E. $\{Y_t, t \in T\}$ es **estacionario en sentido fuerte** cuando la función de distribución conjunta de cualquier vector aleatorio k-dimensional es idéntica para todo subconjunto de índices $\{t_1, ..., t_k\}$ y todo entero $m > 0$:

  $$F(Y_{t_1}, Y_{t_2}, ..., Y_{t_k}) = F(Y_{t_1+m}, Y_{t_2+m}, ..., Y_{t_k+m}) \;\; \forall k, m$$

  En este caso **todos** los momentos de las distribuciones conjuntas de dichos vectores son independientes de los t que los definen.
- La estacionariedad fuerte es importante desde el punto de vista teórico, pero no se requerirá en la práctica y sólo se puede contrastar bajo supuestos distribucionales.

Procesos Estocásticos - Estacionariedad Debil

- Se dice que un P.E. es **estacionario en sentido débil** cuando los momentos de primer y segundo orden de las distribuciones conjuntas de vectores aleatorios son independientes de los t que los definen:
  1. La media es constante: $\mu_t = \mu \;\; \forall t$
  2. La covarianza sólo depende de la distancia entre observaciones: $\gamma_k = Cov(Y_t, Y_{t-k}) = E[(Y_t - \mu)(Y_{t-k} - \mu)] \;\; \forall t, k$
  3. La varianza es constante: $Var(Y_t) = \gamma_0 \;\; \forall t$

  En este caso se tiene:

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} = \frac{\gamma_k}{\gamma_0}$$

- Notar: Con una distribución Normal conjunta, estacionariedad débil implica estacionariedad fuerte, ya que todos sus momentos son función de los dos primeros.

Procesos Estocásticos - Ergodicidad

- La inferencia sobre los momentos de un P.E. es simple cuando éste cumple la propiedad de *ERGODICIDAD* o *DEPENDENCIA DÉBIL*
- Una condición necesaria (no suficiente) para que un P.E. sea **ergódico** es que:

$$\underset{k \longrightarrow \infty}{Lim} \rho_k = 0$$

  Intuitivamente: Si un P.E. muestra 'alta persistencia' significa que valores de la serie muy alejados en el tiempo están altamente correlacionados, lo que hace que cada valor adicional agregue relativamente poca información nueva.
- La no ergodicidad produce problemas de consistencia en la estimación.
- En lo que queda del capítulo supondremos que los procesos son estacionarios y ergódicos salvo que se explicite lo contrario.

Procesos Estocásticos - Estimación de los momentos

- Si un P.E. es débilmente estacionario, sus momentos se pueden estimar a partir de la serie temporal (muestra) utilizando el prinicpio de analogía:

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

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

$$\widehat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k} (Y_t - \widehat{\mu})(Y_{t+k} - \widehat{\mu})$$

- Si un P.E. es además ergódico, las estimaciones anteriores serán consistentes.

PGI y Procesos Estocásticos Lineales

- Estimar los momentos permite aproximar los parámetros del **verdadero P.E.** que habría dado lugar a la serie observada.
  - Los **modelos univariantes** explican Y_t exclusivamente a partir de la información contenida en sus valores pasados (conjunto de información relevante en el **Proceso Generador de Información** -PGI).
- Los **modelos univariantes lineales** representan el P.E. **estacionario** como una combinación **lineal** de variables aleatorias, que pueden ser:
  - retardos de y_t (modelos autorregresivos -AR),
  - retardos de los shocks (modelos de medias móviles -MA),
  - o una combinación de ambos (modelos ARMA).
- El principal objetivo es realizar **predicción**, y los modelos univariantes lineales logran buenos resultados.

Algunos Procesos Estocásticos Lineales Estacionarios

- Proceso Ruido Blanco
- Procesos autorregresivos
  - Modelos AR(1)
  - Modelos AR(p)
- Procesos de medias móviles
  - Modelos MA(1)
  - Modelos MA(q)
- Procesos autorregresivos de medias móviles
  - Modelos ARMA(p,q)

Proceso Ruido Blanco (R.B.)

- Se trata de un proceso puramente aleatorio, $Y_t = \varepsilon_t$, donde:
  - $E[\varepsilon_t] = 0, \forall t$
  - $E[\varepsilon_t]^2 = Var[\varepsilon_t] = \sigma^2, \forall t$
  - $E[\varepsilon_t \varepsilon_s] = Cov[\varepsilon_t, \varepsilon_s] = 0, \forall t, s$ con $t \neq s$
- Si se supone además que su distribución es Normal, se dice que es un 'R.B. Gaussiano'.
- Típicamente, la perturbación de un modelo (en Series Temporales llamada 'shock' o 'innovación') es un proceso R.B.

Modelos Autorregresivos

- Introducidos por Yule (1927) tienen como forma general:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

  - donde ε_t es un proceso Ruido Blanco

- Utilizando un polinomio en el operador de retardos:

$$\Phi(L).Y_t = \delta + \varepsilon_t$$

  - donde se denomina polinomio característico a:

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

Modelos AR(1) - Presentación (1)

- El modelo AR(1) es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \varepsilon_t$$

- Haciendo N sustituciones sucesivas en la primera expresión se tiene:

$$Y_t = \phi_1^{N+1} Y_{t-N-1} + \delta \sum_{j=0}^{N} \phi_1^j + \sum_{j=0}^{N} \phi_1^j \varepsilon_{t-j}$$

- El límite cuando $N \to \infty$ depende críticamente del valor de ϕ_1
- Recordando que con $|\alpha| < 1$ la suma infinita $(1 + \alpha + \alpha^2 + \alpha^3 + ...)$ converge a $\frac{1}{(1-\alpha)}$, con $|\phi_1| < 1$ se tendrá:

$$Y_t = \frac{\delta}{(1 - \phi_1)} + \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$$

Modelos AR(1) - Presentación (2)

- Lo mismo se obtiene en la expresión en el operador de retardos:

$$(1-\phi_1 L)Y_t = \delta + \varepsilon_t$$

$$Y_t = \frac{\delta}{(1-\phi_1 L)} + \frac{\varepsilon_t}{(1-\phi_1 L)}$$

- Con $|\phi_1| < 1$:

$$Y_t = \frac{\delta}{(1-\phi_1 L)} + (1 + \phi_1 L + \phi_1^2 L^2 + \phi_1^3 L^3 + ...)\varepsilon_t$$

$$Y_t = \frac{\delta}{(1-\phi_1)} + \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}$$

Obs: Un proceso AR(1) con $|\phi_1| < 1$ se puede representar como MA(∞)

Modelos AR(1) - Estacionariedad

- $|\phi_1| < 1$ es condición de estacionariedad en un modelo AR(1), ya que:

$$E(Y_t) = \frac{\delta}{(1-\phi_1)} = \mu$$

$$E(Y_t - \mu)^2 = E\left[\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j}\right]^2 = \sigma^2(1 + \phi_1^2 + \phi_1^4 + ...) = \frac{\sigma^2}{(1-\phi_1^2)} = \gamma_0$$

$$E[(Y_t - \mu)(Y_{t-k} - \mu)] = E\left[(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j})(\sum_{j=0}^{\infty} \phi_1^j \varepsilon_{t-j-k})\right] =$$

$$= \sigma^2 \phi_1^k (1 + \phi_1^2 + \phi_1^4 + ...) = \frac{\sigma^2 \phi_1^k}{(1-\phi_1^2)} = \gamma_k$$

  - ¿Qué comportamiento de Y_t cabe esperar $|\phi_1| = 1$?¿Y si $|\phi_1| > 1$?

Modelos AR(1) - Función de Autocorrelación

- A partir de las expresiones obtenidas para la varianza y la covarianza, se obtiene el **coeficiente de correlación para k retardos en un modelo AR(1)**:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k$$

Definición

La **Función de Autocorrelación (FAC)** viene dada por el valor de ρ_k para los distintos valores de k, con $k = 1, 2, ..., \overline{k}$.

- En un AR(1) la FAC Teórica muestra un **decrecimiento exponencial**, más acentuado cuanto menor sea $|\phi_1|$. En particular, se tiene: $\rho_0 = 1, \rho_1 = \phi_1, \rho_2 = \phi_1^2, ..., \rho_{\overline{k}} = \phi_1^{\overline{k}}$

Modelos AR(1) - Ejemplo de FAC (1)

- Ejemplo: '**Correlograma**' para: $Y_t = 0{,}7Y_{t-1} + \varepsilon_t$ con $\varepsilon_t \sim IIN(0, 4)$

Sample: 1 250
Included observations: 250

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |****** |	. |****** |	1	0.725	0.725	132.99	0.000
. |**** |	. |. |	2	0.503	–0.048	197.27	0.000
. |*** |	. |. |	3	0.330	–0.037	225.05	0.000
. |** |	. |. |	4	0.206	–0.016	235.92	0.000
. |* |	. |. |	5	0.115	–0.022	239.33	0.000
. |. |	. |. |	6	0.036	–0.050	239.67	0.000
. |. |	. |. |	7	–0.007	0.004	239.68	0.000
. |. |	. |. |	8	–0.003	0.050	239.68	0.000
. |. |	. |. |	9	–0.017	–0.041	239.75	0.000
* |. |	* |. |	10	–0.060	–0.083	240.71	0.000
* |. |	* |. |	11	–0.110	–0.063	243.91	0.000
. |. |	. |* |	12	–0.040	0.191	244.32	0.000

Figure 14.4 Correlogram of AR(1)

Tomado de Baltagi (2011), 'Econometrics', Springer, Nueva York, pág 376.

Modelos AR(1) - Ejemplo de FAC (2)

- La FAC del proceso AR(1) muestra que la existencia de una estructura autorregresiva de primer orden se expresa en una correlación simple no nula entre Y_t y varios de sus primeros retardos Y_{t-k}.
- La incidencia de Y_{t-k} en Y_t es en este caso exclusivamente indirecta para cualquier $k > 1$ y se da a consecuencia de la correlación ϕ_1 entre los valores intermedios sucesivos Y_{t-k+1} a Y_{t-1}.
- ¿Cómo sería la FAC si $\phi_1 = -0{,}7$?
- ¿y si el modelo fuera $Y_t = 1.Y_{t-1} + \varepsilon_t$ (no estacionario $\Longrightarrow$ no AR(1))?

Modelos AR(1) - Función de Autocorrelación Parcial

- El **coeficiente de autocorrelación parcial** para un retardo k muestra la correlación existente entre Y_t y Y_{t-k} una vez que se depuran los efectos indirectos a través de Y_{t-k+1} a Y_{t-1}.

$$
\begin{aligned}
Y_t &= \delta + \phi_{11} Y_{t-1} + \varepsilon_t \\
Y_t &= \delta + \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + \varepsilon_t \\
&(...) \\
Y_t &= \delta + \phi_{k1} Y_{t-1} + \phi_{k2} Y_{t-2} + ... + \phi_{kk} Y_{t-k} + \varepsilon_t
\end{aligned}
$$

Definición

La **FACP** viene dada por el valor del coeficiente de autocorrelación parcial para los distintos valores de k, con $k = 1, 2, ..., \overline{k}$.

- La FACP toma los mismos valores que la FAC para $k = 1$.
- En un proceso AR(1) el coeficiente de autocorrelación parcial debe ser nulo para cualquier retardo $k > 1$.

Significación de los coeficientes de autocorrelación

- Cuando se conoce la forma del verdadero P.E. puede obtenerse la FAC y la FACP teóricas
- Cuando se trabaja sobre una serie la FAC y la FACP son estimadas a partir de la muestra
- La significación de estos estimadores es relevante:
  - Si $\{Y_t, t \in T\}$ es una secuencia iid con varianza finita, $\rho_k \overset{a}{\sim} N(0, \frac{1}{T})$, pueden construirse bandas de confianza como $\pm 1,96/\sqrt{T}$
  - El estadístico **Q de Ljung-Box** (1978) permite contrastar la significación conjunta de todos los retardos hasta k:

$$H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_k = 0$$
$$H_1 : \rho_j \neq 0 \text{ para algún } j = \{1, 2, ..., k\}$$

$$Q(k) = T(T+2)\sum_{j=1}^{k} \frac{\rho_j^2}{T-j} \sim \chi_k^2$$

Modelos AR(2) - Estacionariedad

- El modelo AR(2) es:

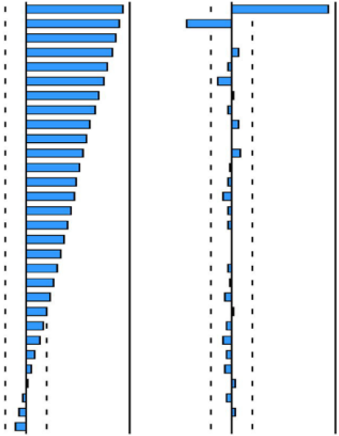
$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad \text{ó} \quad (1 - \phi_1 L - \phi_2 L^2) Y_t = \delta + \varepsilon_t$$

- **Condición de estacionariedad**:
  - raíces del polinomio de retardos $(1 - \phi_1 L - \phi_2 L^2)$ fuera del círculo unitario.
  - raíces de la parte homogénea de Y_t dentro del círculo unitario
  - condición necesaria (no suficiente): $\sum \phi_i < 1$
- Por sustituciones sucesivas u operando con el polinomio de retardos se puede expresar un AR(2) como un MA(∞).

Modelos AR(2) - Correlograma

- Un proceso AR(2) muestra coeficientes de autocorrelación parcial significativamente distintos de cero para los retardos 1 y 2, y coeficientes nulos para retardos mayores.

Sample: 2003:01 2010:03
Included observations: 87



	AC	PAC	Q-stat.	Prob.
1	0.981	0.981	86.718	0.000
2	0.946	−0.456	168.30	0.000
3	0.905	−0.004	243.74	0.000
4	0.863	0.074	313.24	0.000
5	0.823	−0.034	377.16	0.000
6	0.780	−0.138	435.30	0.000
7	0.735	0.010	487.64	0.000
8	0.690	−0.026	534.27	0.000
9	0.646	0.060	575.75	0.000
10	0.607	0.005	612.74	0.000
11	0.572	0.084	646.08	0.000
12	0.541	−0.020	676.33	0.000
13	0.512	−0.032	703.76	0.000
14	0.481	−0.092	728.32	0.000
15	0.448	−0.041	749.90	0.000
16	0.413	−0.041	768.50	0.000
17	0.378	0.004	784.34	0.000
18	0.345	−0.008	797.71	0.000
19	0.312	−0.031	808.81	0.000
20	0.279	−0.018	817.79	0.000
21	0.243	−0.076	824.71	0.000
22	0.206	0.010	829.78	0.000
23	0.169	−0.042	833.25	0.000
24	0.131	−0.096	835.37	0.000
25	0.093	−0.050	836.44	0.000
26	0.051	−0.073	836.78	0.000
27	0.011	0.034	836.79	0.000
28	−0.029	−0.046	836.90	0.000
29	−0.067	0.026	837.50	0.000
30	−0.102	0.007	838.91	0.000

Modelos AR(p)

- El modelo AR(p) es:

$$
\begin{aligned}
Y_t &= \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \\
(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) Y_t &= \delta + \varepsilon_t
\end{aligned}
$$

- **Condición de estacionariedad**: raíces del polinomio de retardos $(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$ fuera del círculo unitario.
- Si Y_t es estacionaria, $E(Y_t) = \mu$, entonces:

$$
\begin{aligned}
\mu &= \delta + \phi_1 \mu + \phi_2 \mu + ... + \phi_p \mu \\
\mu &= \frac{\delta}{1 - \phi_1 - \phi_2 - ... - \phi_p}
\end{aligned}
$$

Modelos AR(p) - Segundos momentos

- Para obtener las covarianzas podemos suponer $\delta = 0$
  - Multiplicando Y_t por Y_{t-k} y tomando esperanzas:

$$\begin{aligned} E(Y_t Y_{t-k}) &= \phi_1 E(Y_{t-1} Y_{t-k}) + \phi_2 E(Y_{t-2} Y_{t-k}) + ... + \\ &+ \phi_p E(Y_{t-p} Y_{t-k}) + E(\varepsilon_t Y_{t-k}) \\ \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + ... + \phi_p \gamma_{k-p} + E(\varepsilon_t Y_{t-k}) \end{aligned}$$

  - Con $k = 0$ se tiene la varianza:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + ... + \phi_p \gamma_p + \sigma^2$$

  - Para $k > 0$:

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + ... + \phi_p \gamma_{k-p} \\ \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + ... + \phi_p \rho_{k-p} \end{aligned}$$

Modelos AR(p) - Ecuaciones de Yule-Walker

- Tomando ρ_k para $k = 1, 2, ..., p$ se obtienen las **Ecuaciones de Yule-Walker**

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + ... + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + ... + \phi_p\rho_{p-2} \\
&(...) \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + ... + \phi_p
\end{aligned}$$

  - La resolución (generalmente matricial) de este sistema, puede permitir obtener los coeficientes de correlación teóricos para procesos conocidos (ϕ_j conocidos).

Modelos AR(p) - Ecuaciones de Yule-Walker (2)

- Análogamente, invirtiendo la matriz de correlaciones y premultiplicando, se pueden utilizar las ecuaciones de Yule-Walker para obtener estimaciones de los coeficientes autorregresivos a partir de los coeficientes de correlación estimados en base a una muestra:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

Modelos de Medias Móviles

- Introducidos por Yule (1921, 1926), explican Y_t como una suma ponderada de los shocks pasados.
- Su forma general es:

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

- Utilizando el operador de retardos:

$$Y_t = \delta + \Theta(L)\varepsilon_t$$

  - con $\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$

Modelos MA(q) - Invertibilidad

- Un proceso MA(q) puede expresarse como un AR(∞) siempre que se cumpla que las raíces de $\Theta(L)$ estén fuera del círculo unitario.
- **Ejemplo**: MA(1) sin constante y con $|\theta_1| < 1$

$$
\begin{aligned}
Y_t &= (1 - \theta_1 L)\varepsilon_t \\
\varepsilon_t &= \frac{1}{(1 - \theta_1 L)} Y_t = \sum_{j=0}^{\infty} (\theta_1 L)^j Y_t = Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \ldots \\
Y_t &= \varepsilon_t - \theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} - \ldots
\end{aligned}
$$

  - La condición para que en el límite se pueda asumir la equivalencia $\frac{1}{(1-\theta_1 L)} \sim \sum_{j=0}^{\infty} (\theta_1 L)^j$ es que la raíz de $(1 - \theta_1 L)$ caiga fuera del círculo unitario:

$$
|L_0| = \left| \frac{1}{\theta_1} \right| > 1 \iff |\theta_1| < 1
$$

Modelos MA(q) - Estacionariedad

- Los procesos MA son por definición **estacionarios** (independientemente de los valores de los θ_j):
  - Su esperanza coincide con la constante del modelo:

$$E(Y_t) = \delta + \Theta(L)E(\varepsilon_t) = \delta$$

  - Para los momentos de segundo orden (asumiendo $\delta = 0$) se multiplica por Y_{t-k} y se toman esperanzas:

$$\begin{aligned} E(Y_t Y_{t-k}) = E\big[&(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q})\cdot \\ &\cdot(\varepsilon_{t-k} - \theta_1\varepsilon_{t-k-1} - \theta_2\varepsilon_{t-k-2} - \ldots - \theta_q\varepsilon_{t-k-q})\big] \end{aligned}$$

$$\Rightarrow \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2$$

$$\Rightarrow \gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q)\sigma^2 & \text{si } k = 1, 2, \ldots, q \\ 0 & \text{si } k > q \end{cases}$$

Modelos MA(q) - Coeficientes de autocorrelación

MA(1)	MA(2)
$\gamma_0 = (1+\theta_1^2)\sigma^2$	$\gamma_0 = (1+\theta_1^2+\theta_2^2)\sigma^2$
$\gamma_1 = -\theta_1\sigma^2$	$\gamma_1 = (-\theta_1+\theta_1\theta_2)\sigma^2$
$\gamma_k = 0, \forall k > 1$	$\gamma_2 = -\theta_2\sigma^2$
	$\gamma_k = 0, \forall k > 2$

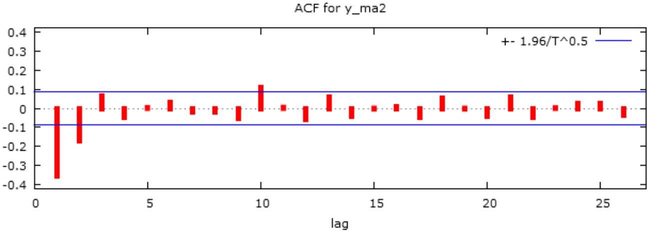
- A partir de la varianza y las autocovarianzas se obtienen los **coeficientes autocorrelación**:

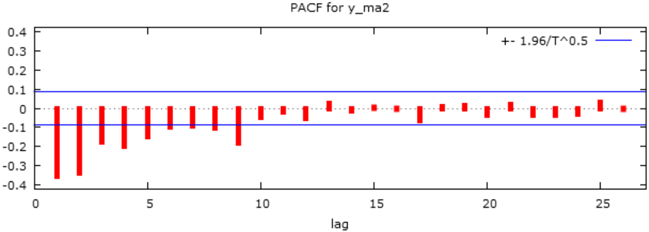
$$\rho_k = \begin{cases} \frac{(-\theta_k+\theta_1\theta_{k+1}+\ldots+\theta_{q-k}\theta_q)}{(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)} & \text{si } k = 1, 2, \ldots, q \\ 0 & \text{si } k > q \end{cases}$$

Modelos MA(2) - Correlograma

- Con una serie de 300 observaciones **simuladas** con distribución N(0,4) se contruyó en Gretl el siguiente proceso MA(2):

$$Y_t = \varepsilon_t - 0,6\varepsilon_{t-1} - 0,3\varepsilon_{t-2}$$





Modelos Autorregresivos de Medias Móviles - ARMA

- Introducidos por Wold (1938) y Barlett (1946), y popularizados por Box y Jenkins (1976)
- Se definen como una combinación de procesos AR y MA
- Forma general de un ARMA(p,q):

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$$

  - O expresado con operadores de retardos:

$$\Phi(L) Y_t = \delta + \Theta(L) \varepsilon_t$$

Modelos ARMA(p,q) - Estacionariedad

- Dado que la parte MA es siempre estacionaria, el modelo ARMA será estacionario cuando la parte AR lo sea
  - **Condición de estacionariedad**: raíces del polinomio de retardos $(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$ fuera del círculo unitario
- Un ARMA(p,q) estacionario puede representarse como MA(∞):

$$Y_t = \frac{\delta}{\Phi(L)} + \frac{\Theta(L)}{\Phi(L)}\varepsilon_t = \frac{\delta}{\Phi(L)} + \Psi(L)\varepsilon_t$$

- La media del proceso será entonces:

$$E(Y_t) = \frac{\delta}{1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p} = \frac{\delta}{1 - \phi_1 - ... - \phi_p}$$

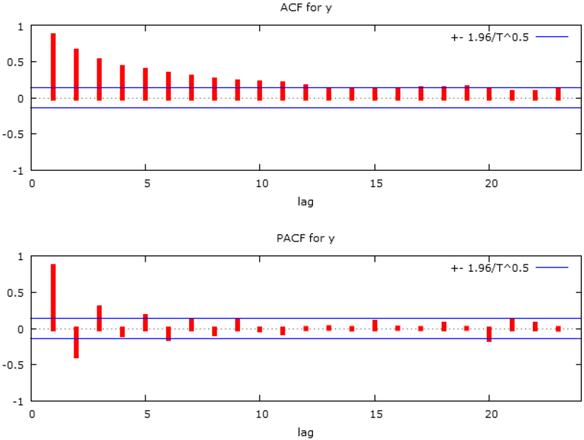
Modelos ARMA(p,q) - Invertibilidad

- El modelo ARMA será invertible cuando la parte MA lo sea:
  - **Condición de invertibilidad**: raíces del polinomio de retardos $(1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q)$ fuera del círculo unitario
- Un ARMA(p,q) invertible puede representarse como AR(∞):

$$\frac{\Phi(L)}{\Theta(L)} Y_t = \Pi(L) Y_t = \frac{\delta}{\Theta(L)} + \varepsilon_t$$

Modelos ARMA(p,q) - FAC y FACP

- Si se tiene un modelo ARMA(p,q) estacionario e invertible:
  - Por representación AR: la FAC decrece lentamente (en val. abs.)
  - Por representación MA: la FACP decrece lentamente (en val. abs.)
  - Simulando un proceso ARMA(1,1): $Y_t = 0,75Y_{t-1} + \varepsilon_t + 0,8\varepsilon_{t-1}$ a partir de 250 observaciones de $\varepsilon_t \sim N(0,4)$:



Modelos ARMA(1,1)

- Forma general:

$$\begin{aligned} Y_t &= \delta + \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ (1 - \phi_1 L) Y_t &= \delta + (1 - \theta_1 L)\varepsilon_t \end{aligned}$$

- En este caso en su representación MA(∞):

$$\Psi(L) = \frac{\Theta(L)}{\Phi(L)} \Leftrightarrow \Phi(L)\Psi(L) = \Theta(L)$$

  - Igualando término a término en:

$$(1 - \phi_1 L)(1 + \psi_1 L + \psi_2 L^2 + ...) = (1 - \theta_1 L)$$

  - Se tiene:

$$[1 + (\psi_1 - \phi_1)L + (\psi_2 - \phi_1\psi_1)L^2 + (\psi_3 - \phi_1\psi_2)L^3 + ...] = (1 - \theta_1 L)$$

Modelos ARMA(1,1)

- Obtención de los coeficientes del MA(∞):

$$\begin{array}{lcl} \psi_1 - \phi_1 = -\theta_1 & \Rightarrow & \psi_1 = \phi_1 - \theta_1 \\ \psi_2 - \phi_1 \psi_1 = 0 & \Rightarrow & \psi_2 = \phi_1 \psi_1 \\ \psi_k - \phi_1 \psi_{k-1} = 0 & \Rightarrow & \psi_k = \phi_1 \psi_{k-1} \end{array}$$

- Para procesos ARMA(p,q) se obtienen soluciones similares, donde los q primeros coeficientes ψ se ven afectados por los θ, y los coeficientes ψ_k para $k > q$ se obtienen como función de los valores ψ_{k-1} a ψ_{k-p}
- En forma análoga se obtienen los coeficientes de $\Pi(L)$ de la representación AR(∞) de un ARMA(1,1) o, más en general, de un ARMA(p,q)

Modelos ARMA(1,1) - Segundos momentos

- Tomando el ARMA(1,1) sin constante:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

- Multiplicando por Y_{t-k} y tomando esperanzas:

$$\gamma_k = E(Y_t Y_{t-k}) = \phi_1 \gamma_{k-1} + E(\varepsilon_t Y_{t-k}) - \theta_1 E(\varepsilon_{t-1} Y_{t-k})$$

  - A partir de esta expresión se derivan las autocovarianzas y con ellas se obtienen los coeficientes de autocorrelación (ver Uriel, pág 56):

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - 2\phi_1\theta_1 + \theta_1^2}$$
$$\rho_k = \phi_1 \rho_{k-1} \; \forall k > 1$$

Modelos ARMA - Resumen FAC y FACP

Características teóricas de la FAC y de la FAP de los procesos estacionarios

Proceso	**FAC:** $\rho_1, \ldots, \rho_k$	**FACP:** $\phi_{11}, \ldots, \phi_{kk}$
R.B.	Todos los $\rho_k = 0$	$\phi_{kk} = 0$
AR(1)	Decae exponencialmente hacia cero, en forma directa [si $\phi_1 > 0$] o alternando signos [si $\phi_1 < 0$]. $\rho_k = \phi_1^k$	$\phi_{11} = \rho_1 = \phi_1$ $\phi_{kk} = 0$ para $k \geq 2$
AR(p)	Decae hacia cero, los coeficientes pueden oscilar	$\phi_{kk} \neq 0$ para $k \leq p$ $\phi_{kk} = 0$ para $k > p$
MA(1)	$\rho_1 < 0$ [si $\theta_1 > 0$] o $\rho_1 > 0$ [si $\theta_1 < 0$]	Decae directamente [si $\theta_1 < 0$] o alternando signos [si $\theta_1 > 0$]. $Sg(\phi_{11}) = Sg(\theta_1)$
ARMA(1,1)	Decae exponencialmente hacia cero a partir del retardo 1, directamente [si $\phi_1 > 0$] o alternando signos [si $\phi_1 < 0$]. $Sg(\rho_1) = Sg(\phi_1 - \theta_1)$	Decae exponencialmente hacia cero a partir del retardo 1, alternando signos [si $\phi_1 > 0$] o directamente [si $\phi_1 < 0$]. $\phi_{11} = \rho_1$. $Sg(\rho_1) = Sg(\phi_1 - \theta_1)$
ARMA(p,q)	Decae directa o exponencialmente hacia cero a partir del retardo *q*	Decae directa o exponencialmente hacia cero a partir del retardo *p*

Basado en Enders (1995), Tabla 2.1.

Procesos Integrados

- Los modelos ARMA aplican exclusivamente a series estacionarias
- Hay diferentes tipos de procesos estocásticos **no estacionarios**
  - Ej: raíces del polinomio de retardos dentro del círculo unitario
    - En este caso los procesos serán **explosivos**
    - Ejemplo: proceso autorregresivo con $|\phi| > 1$
- En economía interesan especialmente los procesos que llamamos de **raíz unitaria**
  - El polinomio de retardos tiene al menos una raíz igual a 1
  - En este caso el proceso absorbe cada shock en forma permanente, tiene **memoria infinita**
  - La FAC decae en forma extremadamente lenta

Ejemplo: Paseo Aleatorio

- Caso simple de proceso integrado: **Paseo Aleatorio** o Random Walk

$$Y_t = Y_{t-1} + \varepsilon_t \quad \text{ó} \quad \Delta Y_t = \varepsilon_t$$

  - Haciendo t sustituciones sucesivas y suponiendo una condición inicial:

$$Y_t = Y_0 + \sum_{j=1}^{t} \varepsilon_j \qquad \text{donde } E(Y_t) = E(Y_{t-s}) = Y_0$$

  - O si el proceso se inicia en un pasado infinitamente remoto:

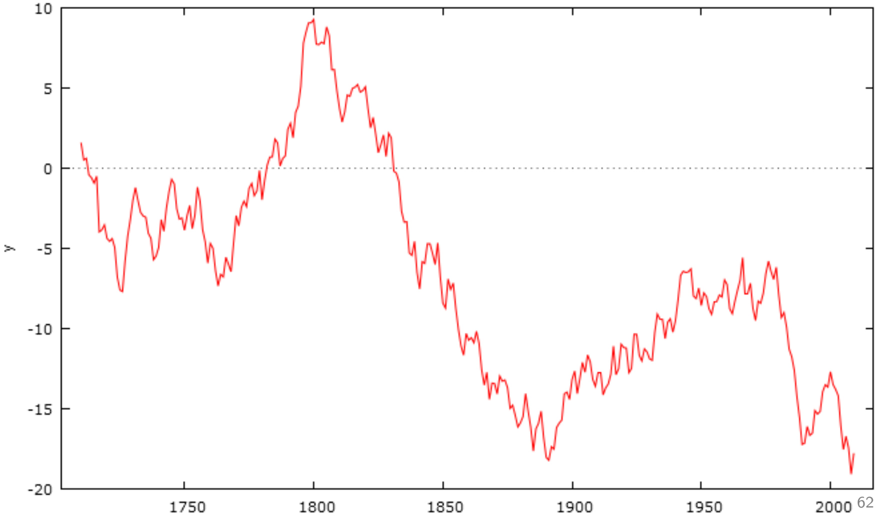
$$Y_t = \sum_{j=0}^{\infty} \varepsilon_{t-j} \qquad \text{donde } E(Y_t) = E(Y_{t-s}) = 0$$

  - Pero, el proceso es **no estacionario** por tener varianza infinita:

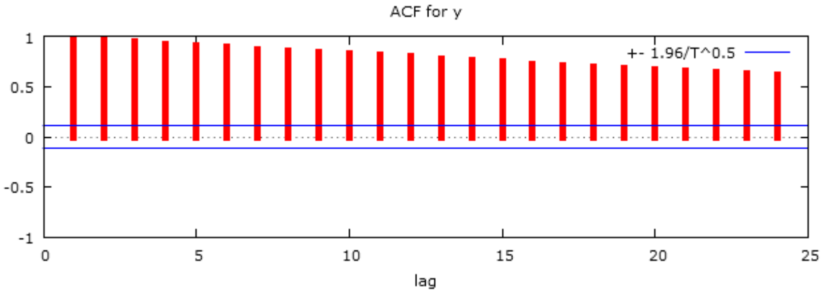
$$Var(Y_t) = Var(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + ...) = t\sigma^2$$

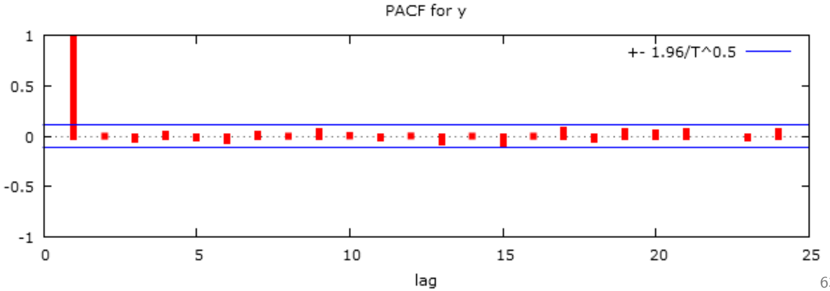
Paseo Aleatorio: Ejemplo

- Ejemplo de paseo aleatorio: $Y_t = Y_{t-1} + \varepsilon_t$
  - Simulación con 300 realizaciones pseudoaleatorias de un RBG: $\varepsilon_t \sim N(0, 1)$



Paseo Aleatorio: Ejemplo





Tendencia Estocástica

- Una **tendencia determinística** proviene de que el nivel de la serie incorpora en forma permanente un componente determinista (p.e. constante si la tendencia es lineal)
- La **tendencia estocástica** proviene de la incorporación en forma permanente de cada shock, que se da en todos los procesos integrados
  - Característica fundamental: la serie tiende a **deambular**, no muestra **reversión a la media**
  - En un período corto, una tendencia estocástica puede confundirse con una tendencia determinística
- Frecuentemente las series económicas presentan a la vez tendencia estocástica y tendencia determinística
  - Una tarea crítica en este análisis será discernir la forma adecuada para aproximar los comportamientos tendenciales observados

Paseo Aleatorio con deriva ('drift')

- Cuando en un paseo aleatorio se incluye una constante:

$$Y_t = \delta + Y_{t-1} + \varepsilon_t \quad \text{ó} \quad \Delta Y_t = \delta + \varepsilon_t$$

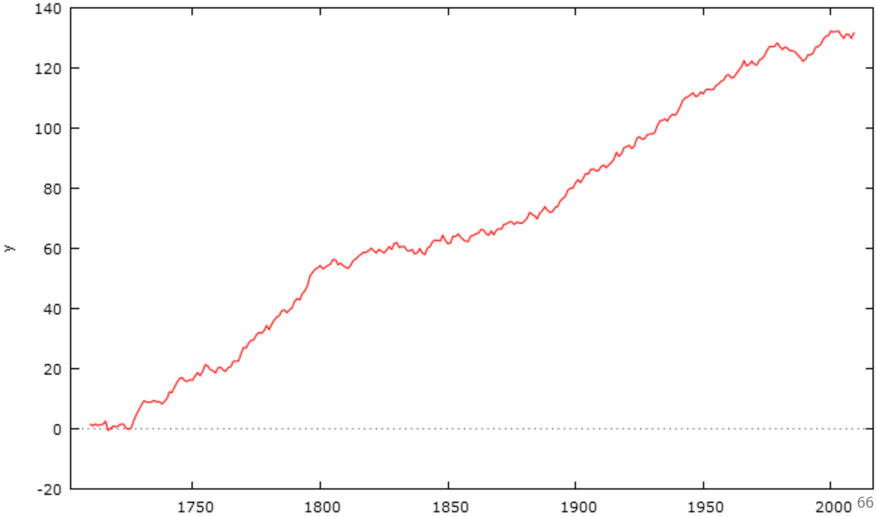
  - La constante se incorpora íntegramente en cada período, por lo que se aprecia como una tendencia determinística lineal
  - Por ejemplo, con condición inicial Y_0:

$$Y_t = Y_0 + \delta t + \sum_{j=1}^{t} \varepsilon_j$$

  - En este caso la serie presenta una tendencia estocástica **y** una tendencia determinística, deambulando alrededor de esta última

Paseo Aleatorio con deriva: Ejemplo

- Ejemplo de paseo aleatorio: $Y_t = 0,5 + Y_{t-1} + \varepsilon_t$
  - Simulación las mismas 300 realizaciones pseudoaleatorias de un RBG: $\varepsilon_t \sim N(0,1)$



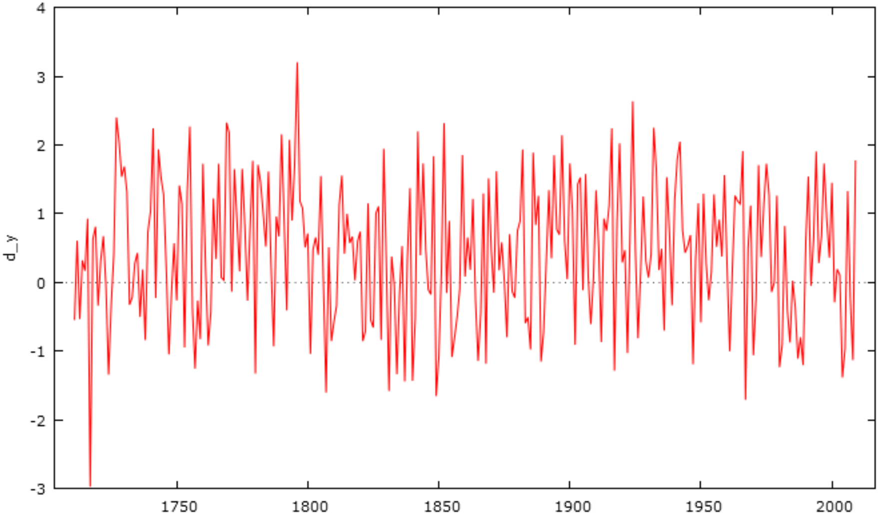
Tendencias y diferenciación

- Si solo existe tendencia determinística
  - El proceso es no estacionario en media, pero estacionario en torno a la tendencia ('**trend stationary**' - TS)
  - Se puede solucionar incluyendo una tendencia en el modelo o quitándole la tendencia a la serie
- Cuando existe una tendencia estocástica
  - La diferenciación es la única forma de eliminar la raíz unitaria ('**difference stationary**' - DS)
  - Además la diferencia elimina la tendencia determinística
  - En el paseo aleatorio con deriva

$$\begin{aligned} Y_t &= \delta + Y_{t-1} + \varepsilon_t \\ \Delta Y_t &= \delta + \varepsilon_t \end{aligned}$$

Diferenciación en el Paseo Aleatorio

- **Ejemplo**: Primera diferencia en el paseo aleatorio con deriva



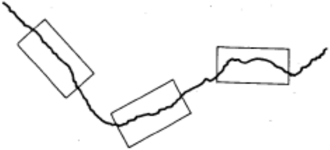
Diferenciación como transformación estacionaria (1)

- Los procesos integrados se pueden transformar en estacionarios a través de la diferenciación
  - El polinomio de retardos $\Phi_p(L)$ puede tener una o más raíces unitarias
  - La transformación estacionaria del proceso requerirá tantas diferenciaciones como raíces unitarias existan
  - Cada vez que se diferencia se elimina UNA raíz unitaria
    - La diferencia de segundo orden es la diferencia de la diferencia: $\Delta(\Delta Y_t)$
    - La diferencia de orden d de Y_t es: $\Delta^d Y_t = (1-L)^d Y_t$
- d es el **orden de integración** del proceso, que se dirá que es $\mathrm{I}(d)$ si d es el mínimo número de diferencias necesarias para transformar el proceso en estacionario
  - Generalmente las series económicas son I(1) o a lo sumo I(2)

Diferenciación como transformación estacionaria (2)

- El gráfico puede ser indicativo del orden de diferenciación:

a) Serie no estacionaria en nivel ($d = 1$)



b) Serie no estacionaria en nivel ni en pendiente ($d = 2$)

- Un riesgo importante es el de **sobrediferenciar** la serie:
  1. Incrementaría la varianza del proceso
  2. generaría estructuras artificiales de medias móviles con raíces $\lesssim 1$
  3. Llevaría a un modelo menos parsimonioso

Diferenciación: ARIMA(p,d,q)

- Un proceso integrado Y_t se denomina **ARIMA(p,d,q)** si tomando diferencias de orden d se lo puede transformar en ARMA(p,q)
  - Entonces, si $\varepsilon_t \sim N(0, \sigma^2)$ un ARIMA(p,d,q) es:

$$\Delta^d Y_t = \delta + \phi_1 \Delta^d Y_{t-1} + ... + \phi_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$$

$$\Phi_p(L)(1-L)^d Y_t = \delta + \Theta_q(L)\varepsilon_t$$
$$(1 - \phi_1 L - ... - \phi_p L^p)(1-L)^d Y_t = \delta + (1 - \theta_1 L - ... - \theta_q L^q)\varepsilon_t$$

- Una serie I(d) posee d raíces unitarias en su polinomio de retardos
  - Si ΔY_t es ARMA(p,q) entonces Y_t es ARIMA(p,1,q)
  - Si ΔY_t es no estacionario y $\Delta^2 Y_t$ es ARMA(p,q), entonces Y_t es ARIMA (p,2,q)
  - Más adelante se verán pruebas para contrastar la existencia de raíces unitarias. En esta metodología frecuentemente se evalúa la estacionariedad en base al correlograma

Metodología de Box-Jenkins

- Un **proceso específico** o **estructura** ARIMA(p,d,q) es $\Phi_p(L)(1-L)^d Y_t = \delta + \Theta_q(L)\varepsilon_t$ una vez que se determinaron los parámetros $d, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \delta, \sigma^2$
- Box y Jenkins (1976) presentan un procedimiento para la inferencia de una estructura ARIMA(p,d,q) a partir de una serie de realizaciones $\{Y_1, Y_2, ... Y_T\}$
- La metodología consta de tres etapas:
  - **Identificación** (o especificación inicial)
  - **Estimación**
  - **Validación**

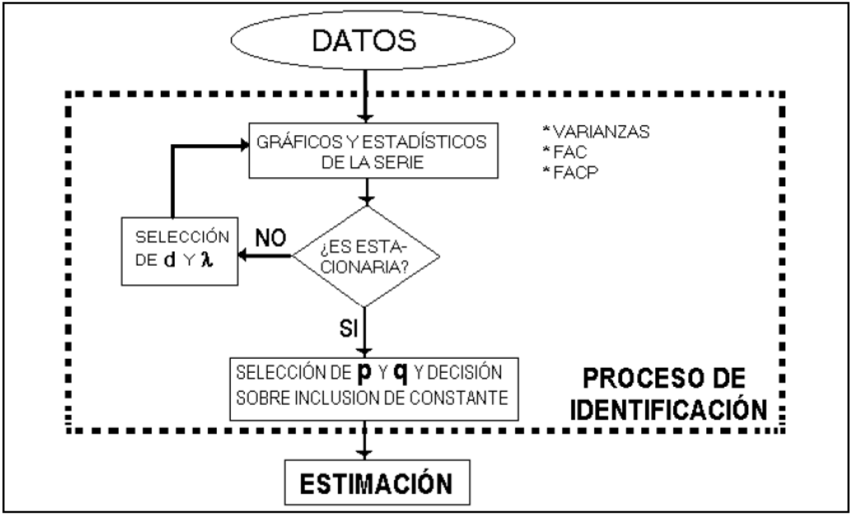
Box-Jenkins: Identificación (1)

- **Objetivos:**
  1. Seleccionar una **transformación estacionaria**
     - Transformaciones para la no estacionariedad en varianza: Box-Cox(λ) (equivalente al Ln si $\lambda = 0$, o a exponentes fraccionarios si $\lambda \leq 1$)
     - Número de diferencias d

$$w_t = (1 - L)^d Y_t^{(\lambda)}$$

  2. Seleccionar los **órdenes de retardos** p y q, y definir si corresponde incluir constante

Box-Jenkins: Identificación (2)



Tomado de Fernández A., 'Análisis de Series de Tiempo', Notas de Clase, FCEyA, Cap 8.

Box-Jenkins: Identificación (3)

- **Instrumentos:**
  - Para la determinación de d
    - Análisis gráfico de la serie y de sus diferencias
    - Análisis de la varianza de la serie y de sus diferencias
    - Análisis del correlograma de la serie y de sus diferencias
    - Contrastes de Raíz Unitaria (ver Tema 6)
  - Para la determinación de p y q
    - Análisis del correlograma de la transformación estacionaria
  - Para determinar si el proceso tiene constante δ
    - Lo más simple es estimar el modelo con constante y luego contrastar su significación
- La identificación en general es tentativa y sugiere más de un modelo posible

Box-Jenkins: Estimación (1)

- La estimación de un modelo ARIMA se realiza estimando los parámetros $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \delta$ del modelo ARMA para la transformación estacionaria e invertible (w_t) a partir de una muestra de tamaño $T - d$ (ya que con cada diferencia se pierde una observación)
- Como es habitual, el análisis de la significación de los coeficientes conduce a una reespecificación, que en este caso implica un retorno a la fase de identificación.
- Aunque la estimación puede hacerse por MCO, al ser una estimación no lineal (si $q \geq 1$) lo más habitual es realizarla por **Máxima Verosimilitud**, asumiendo una distribución condicional concreta $f(\cdot)$, generalmente Normal

Box-Jenkins: Estimación (2)

- La **función de verosimilitud**, considerando la estructura dinámica del proceso estocástico estacionario $\{W_t, t = 1, ..., T\}$, es:

$$L = f(\mathbf{W}_T) = f(w_T|\mathbf{W}_{T-1})f(\mathbf{W}_{T-1}) =$$
$$= f(w_T|\mathbf{W}_{T-1})f(w_{T-1}|\mathbf{W}_{T-2})f(\mathbf{W}_{T-2}) = ...$$

$$L = \prod_{t=1}^{T} f(w_t|\mathbf{W}_{t-1})f(\mathbf{w}^0)$$

$$\log L = \sum_{t=2}^{T} \log f(w_t|\mathbf{W}_{t-1}) + \log f(\mathbf{w}^0)$$

- La **log-verosimilitud** depende de las condiciones iniciales
  - Si se determinan en el proceso de estimación $\rightarrow$ MV Exacta
  - Si se asignan valores iniciales $\mathbf{w}^0$ y **”**^{0} dados $\rightarrow$ MV Condicional (equivalente a MCO)

Box-Jenkins: Estimación (3)

- Si w_t es **condicionalmente Normal**, su densidad condicional es:

$$f(w_t|\mathbf{W}_{t-1}) = f(\varepsilon_t|\sigma^2, \beta, w^0, \varepsilon^0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_t^2}{2\sigma^2}}$$

- donde $\beta = (\phi_1 ... \phi_p \ \theta_1 ... \theta_q)'$, $w^0 = (w_0 ... w_{-p+1})$, y $\varepsilon^0 = (\varepsilon_0 ... \varepsilon_{-q+1})$
- Dado la incorrelación entre los ε_t la densidad conjunta será:

$$f(\varepsilon_1 ... \varepsilon_T|\sigma^2, \beta, w^0, \varepsilon^0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum \varepsilon_t^2}$$

- Considerando que
$\varepsilon_t = w_t - \phi_1 w_{t-1} - ... - \phi_p w_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$, β se obtiene maximizando:

$$f(\varepsilon_1 ... \varepsilon_T|\sigma^2, \beta, w^0, \varepsilon^0) = \frac{e^{-\frac{1}{2\sigma^2}\sum(w_t - \phi_1 w_{t-1} - ... - \phi_p w_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q})^2}}{\sqrt{2\pi\sigma^2}}$$

Box-Jenkins: Validación (1)

- *Resultados de la Estimación*
  - **Significación** de los coeficientes (y vuelta a identficiación)
  - **Estacionariedad** e **Invertibilidad**
  - **Raíces comunes**
- *Análisis de los residuos* $\hat{\varepsilon}_t$
  - **Aproximación a Ruido Blanco**: Gráfico, correlograma, Q de Ljung-Box
  - **Aproximadamente Normalidad**: Histograma, contraste Jarque-Bera (Ver Tema 5)
- *Comparación de modelos*
  - Principio de **parsimonia**
  - Error estándar de la regresión
  - Criterios de información: AIK, BIC y Hannan-Quinn

Box-Jenkins: Validación (2)

- En esta etapa:
  - Se valida o se descarta cada uno de los modelos estimados
  - Se comparan entre ellos los distintos modelos validados
- Los distintos resultados obtenidos pueden sugerir un retorno a las etapas de identificación y estimación
  - Típicamente por el análisis de los resultados de la estimación, pero también cuando el análisis de los residuos revela una incorrecta especificación dinámica
- La comparación de modelos se realiza siguiendo el principio de **parsimonia**, que indica como preferible a un modelo con menor cantidad de parámetros
  - Un ejemplo extremo es que se prefiere la representación AR(1) de un proceso frente a su representación MA(∞)
  - Este principio se recoje también en los criterios de información

Box-Jenkins: Validación (3)

- Los **criterios de información** ayudan a comparar distintos modelos a partir de la log-verosimilitud, penalizando por la inclusión de variables
  - Criterio de Información de **Akaike**:

$$AIC : c_n(k) = \frac{-2Ln(L_n(k))}{n} + \frac{2k}{n}$$

  - Criterio de Información de **Schwarz**:

$$BIC : c_n(k) = \frac{-2Ln(L_n(k))}{n} + \frac{kLn(n)}{n}$$

  - Criterio de Información de **Hannan-Quinn**:

$$HQ : c_n(k) = \frac{-2Ln(L_n(k))}{n} + \frac{2kLn(Ln(n))}{n}$$

- Con cada uno de estos criterios se considera que un modelo es preferible si presenta un valor menor (en general son negativos)

Box-Jenkins: Utilización (1)

- Como resultado de las tres etapas anteriores, se cuenta con un modelo estimado compatible con la estructura de los datos
- Si bien puede haber otros usos, estos modelos tienen como principal objetivo la **predicción**,
  - Se busca pronosticar en términos probabilísticos los valores futuros de una serie temporal
  - Desde los años 80 esta metodología es la utilizada por excelencia para la predicción en distintas áreas
- Además, la **predicción dentro de la muestra** se utiliza como prueba de la bondad de ajuste de un modelo
  - Una discrepancia sistemática entre la predicción y lo observado cuestiona el ajuste

Box-Jenkins: Utilización (2)

- En la **predicción fuera de la muestra** se dispone de información de la variable hasta el momento T y se desea predecir Y_{T+l} ($l > 0$)
- Se vio que un proceso ARMA (estacionario e invertible) puede expresarse como MA(∞):

$$Y_t = \frac{\Theta_q(L)}{\Phi_p(L)}\varepsilon_t = \Psi_\infty(L)\varepsilon_t$$

- Se define el **predictor óptimo** $\widetilde{Y}_{T+l/T}$, que se determina minimizando la varianza del error de predicción a través de la minimización del Error Cuadrático Medio:

$$ECM[\widetilde{Y}_{T+l/T}] = E[Y_{T+1} - \widetilde{Y}_{T+l/T}]^2$$

- El predictor óptimo para el proceso ARMA será:

$$\widetilde{Y}_{T+l/T} = \sum_{j=0}^{\infty} \psi_{l+j}\varepsilon_{t-j}$$

Box-Jenkins: Utilización (3)

- Para el cálculo de las **predicciones** se retoma la formulación ARMA:

$$\widetilde{Y}_{T+l/T} = \varphi_1 \widetilde{Y}_{T+l-1/T} + \ldots + \varphi_{p+d} \widetilde{Y}_{T+l-p-d/T} + \tilde{\varepsilon}_{T+l} - \theta_1 \tilde{\varepsilon}_{T+l-1} - \ldots - \theta_q \tilde{\varepsilon}_{T+l-q}$$

  - El modelo ARMA se aplica recursivamente, obteniendo $\widetilde{Y}_{T+1/T}$, $\widetilde{Y}_{T+2/T}$, ..., $\widetilde{Y}_{T+l/-1T}$, $\widetilde{Y}_{T+l/T}$
    - ★ Para períodos dentro de la muestra se utilizan como $\widetilde{Y}_{T+l-j/T}$ y $\tilde{\varepsilon}_{T+l-j}$ los valores observados
    - ★ Fuera de la muestra se utilizan como $\widetilde{Y}_{T+l-j/T}$ los valores ya predichos y 0 en lugar de $\tilde{\varepsilon}_{T+l-j}$
  - Por ejemplo para un MA(2):

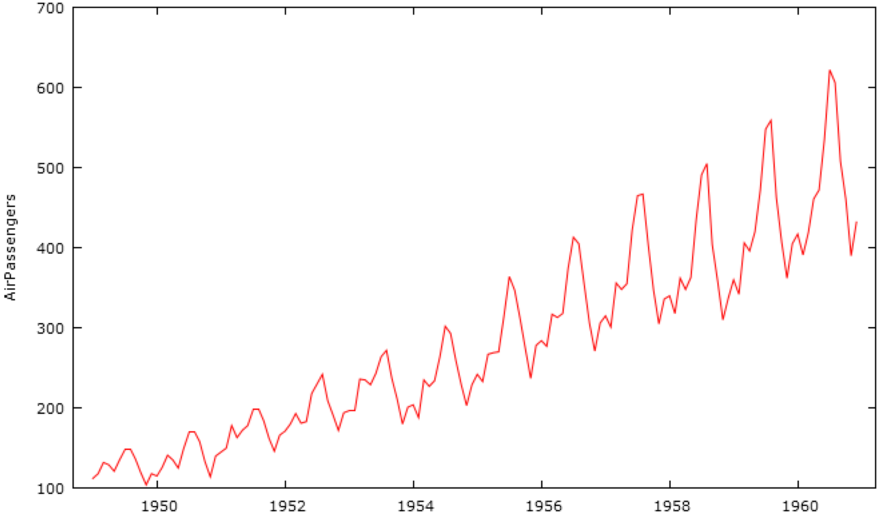
Período	Valor Real	Predictor	ECM
$T+1$	$Y_{T+1} = \varepsilon_{T+1} - \theta_1 \varepsilon_T - \theta_2 \varepsilon_{T-1}$	$\tilde{Y}_{T+1/T} = -\theta_1 \varepsilon_T - \theta_2 \varepsilon_{T-1}$	σ_ε^2
$T+2$	$Y_{T+2} = \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1} - \theta_2 \varepsilon_T$	$\tilde{Y}_{T+2/T} = -\theta_2 \varepsilon_T$	$(1+\theta_1^2)\sigma_\varepsilon^2$
$T+3$	$Y_{T+3} = \varepsilon_{T+3} - \theta_1 \varepsilon_{T+2} - \theta_2 \varepsilon_{T+1}$	$\tilde{Y}_{T+2/T} = 0$	$(1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2$

Estacionalidad - Introducción (1)

- Una serie es **estacional** cuando su evolución muestra un patrón intra-anual sostenido a lo largo de los años
- Se interpreta que la serie está como compuesta por una **parte regular** y una **parte estacional**
- Se denomina s al **período estacional**
  - $s = 4$ si los datos son trimestrales,
  - $s = 12$ si los datos son mensuales

Estacionalidad - Introducción (2)

- **Ejemplo:** Número mensual de pasajeros que viajan en una aerolínea (1949:1-1960:12), tomada de Box y Jenkins (1976)



Estacionalidad - Introducción (3)

- La parte estacional puede evolucionar en forma **determinista**, y en ese caso se elimina o modela con dummies estacionales
- En general la parte estacional se comporta como un **proceso estocástico** en sí mismo
    - Puede ser **no estacionario**
        - Deben aplicarse **diferencias estacionales** hasta obtener su transformación estacionaria
    - Puede ser (o haber sido transformado en) **estacionario**
        - Existen distintas posibilidades para modelarlo como un **ARMA estacional**

Modelos Estacionarios Estacionales Puros

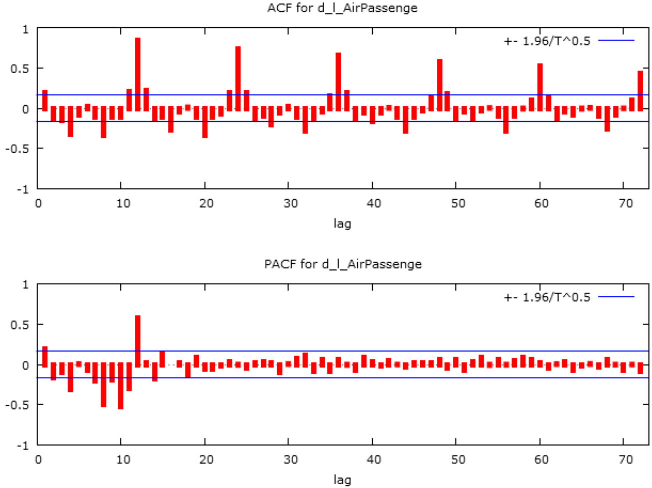
- Una serie estacionaria es **estacional pura** cuando la correlación serial sólo es significativa entre observaciones distanciadas s períodos (no hay parte regular)
- Se interpreta como si fueran s procesos estocásticos independientes, todos con una misma estructura $ARMA$
  - un modelo $AR(1)_s$ será: $Y_t = \phi_1^s Y_{t-s} + \varepsilon_t$.
  - un modelo $MA(2)_s$ será: $Y_t = \varepsilon_t - \theta_1^s \varepsilon_{t-s} - \theta_2^s \varepsilon_{t-2s}$
  - un modelo $ARMA(1,1)_s$ será: $Y_t = \phi_1^s Y_{t-s} + \varepsilon_t - \theta_1^s \varepsilon_{t-s}$
  - un modelo $ARMA(P,Q)_s$ será: $\Phi_P^s(L) Y_t = \Theta_Q^s(L) \varepsilon_t$

Estacionalidad - Correlogramas (1)

- Los **correlogramas teóricos** de estos procesos se comportan igual que los de la parte regular,
  - los coeficientes estacionales aparecen en los retardos s, $2s$, $3s$,...
  - los coeficientes intermedios son nulos
- En los **correlogramas estimados** para series económicas suelen encontrarse patrones estacionales no puros, donde hay simultáneamente un ARMA en la parte regular y un ARMA en la parte estacional
  - la existencia de un fuerte patrón estacional tiende a ocultar las señas de la parte regular
  - en este caso las dummies no elminan adecuadamente la parte estacional (tampoco otros métodos utilizados para el 'ajuste estacional')

Estacionalidad - Correlogramas (2)

- **Ejemplo:** Correlograma de los datos de aerolíneas (primera diferencia regular del logaritmo: dif-log)
  - Se observa la no estacionariedad en la parte estacional



Modelos Estacionarios Estacionales Aditivos

- Una serie estacionaria puede obedecer a un modelo ARMA en su parte regular y otro en su parte estacional
  - Estos procesos muestran asociación tanto entre las observaciones sucesivas como entre las observaciones distanciadas s períodos
- Cuando estas dos estructuras ARMA simplemente coexisten pero no interactúan, se dice que la estacionalidad es **aditiva**
  - Un ejemplo puede ser el siguiente modelo $ARMA(1,1) + ARMA(1,0)_{12}$:

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

$$(1 - \phi_1 L - \phi_{12} L^{12}) Y_t = (1 - \theta_1 L)\varepsilon_t$$

Este modelo es un $ARMA(12,1)$ restringido a que $\phi_2 = ... = \phi_{11} = 0$

Modelos Estacionarios Estacionales Multiplicativos (1)

- En el caso en que haya una interacción entre ambas partes en la estructura de correlación, la estacionalidad será **multiplicativa**
- Un modelo multiplicativo $ARMA(p, q) \times ARMA(P, Q)_s$ es:

$$\Phi_p(L)\Phi_P^s(L)Y_t = \Theta_q(L)\Theta_Q^s(L)\varepsilon_t$$

Modelos Estacionarios Estacionales Multiplicativos (2)

- Para el caso de un $ARMA(1,1) \times ARMA(1,1)_{12}$:

$$(1-\phi_1 L)(1-\phi_1^s L^{12})Y_t = (1-\theta_1 L)(1-\theta_1^s L^{12})\varepsilon_t$$

$$Y_t - \phi_1 Y_{t-1} - \phi_1^s Y_{t-12} + \phi_1 \phi_1^s Y_{t-13} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_1^s \varepsilon_{t-12} + \theta_1 \theta_1^s \varepsilon_{t-13}$$

  Puede expresarse utilizando solamente parámetros no estacionales:

$$Y_t - \phi_1 Y_{t-1} - \phi_{12} Y_{t-12} + \phi_{13} Y_{t-13} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_{12} \varepsilon_{t-12} + \theta_{13} \varepsilon_{t-13}$$

  Notar que es un $ARMA(13,13)$ restringido a $\phi_2 = \ldots = \phi_{11} = \theta_2 = \ldots = \theta_{11} = 0$, y a que se respete la estructura multiplicativa $\phi_{13} = \phi_1 \phi_{12}$ y $\theta_{13} = \theta_1 \theta_{12}$

- El modelo multiplicativo es más parsimonioso, pero su ajuste dependerá de que los datos avalen las restricciones que impone

Procesos Estacionales No Estacionarios

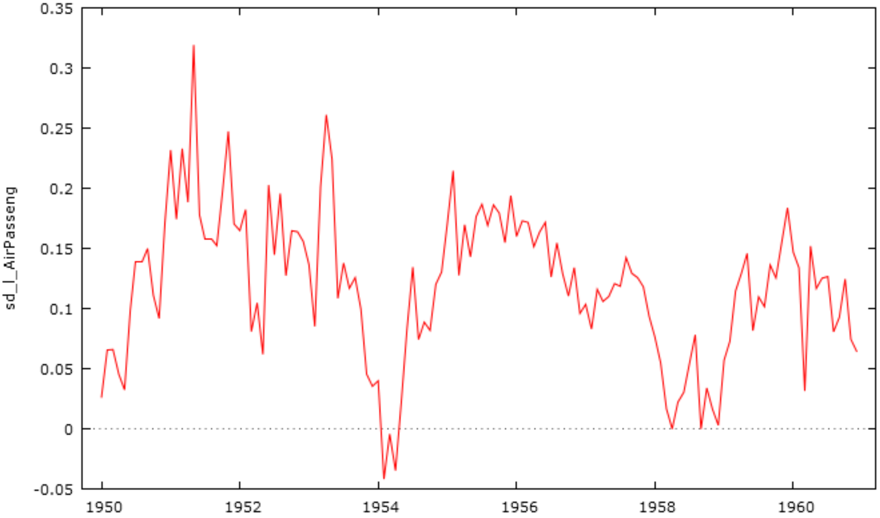
- Para una serie estacional debe establecerse el orden de diferencias necesarias para transformar en estacionaria la parte estacional, al que se nota D
- De este modo, el modelo $ARIMA(p, d, q) \times ARIMA(P, D, Q)_s$ es:

$$\Phi_p(L)\Phi_P^s(L)(1-L)^d(1-L)^D Y_t = \delta + \Theta_q(L)\Theta_Q^s(L)\varepsilon_t$$

- Denominados $SARIMA(p, d, q)(P, D, Q)_s$, son la versión más general de la familia de modelos Box-Jenkins
- La identificación implicará definir el orden de retardos P y Q de los componentes AR y MA de la parte estacional
- Para ello se emplea la misma lógica que la presentada para la parte regular

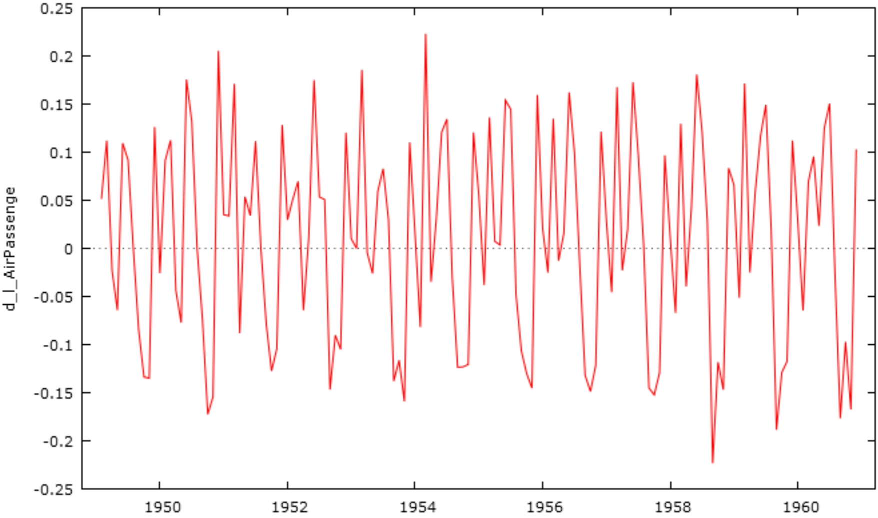
Diferencia estacional y diferencia regular (1)

- **Ejemplo de las Aeorlíneas:** Gráfico de $(1 - L^{12})LnY_t$



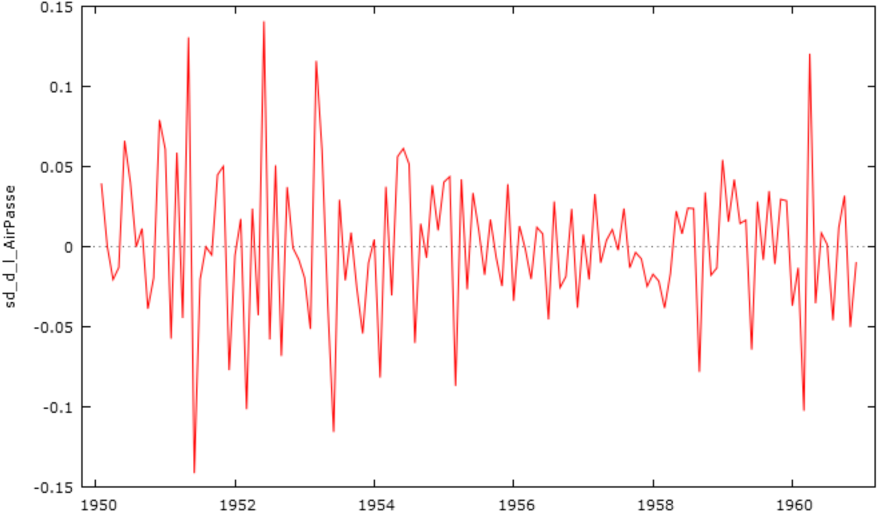
Diferencia estacional y diferencia regular (2)

- **Ejemplo de las Aeorlíneas:** Gráfico de $(1 - L)LnY_t$



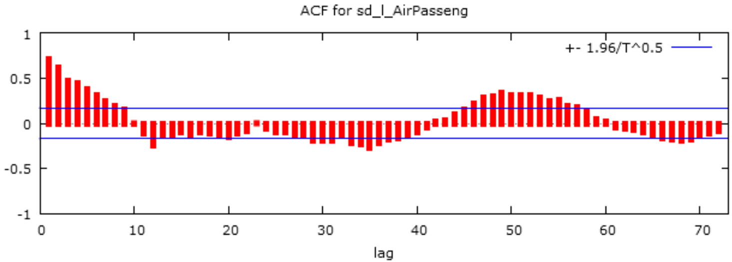
Diferencia estacional y diferencia regular (3)

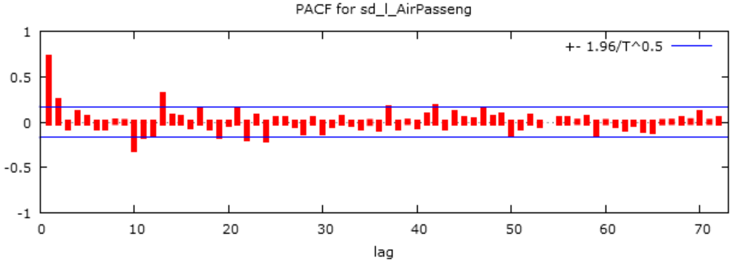
- **Ejemplo de las Aeorlíneas:** Gráfico de $(1-L)(1-L^{12})LnY_t$



Diferencia estacional y diferencia regular (4)

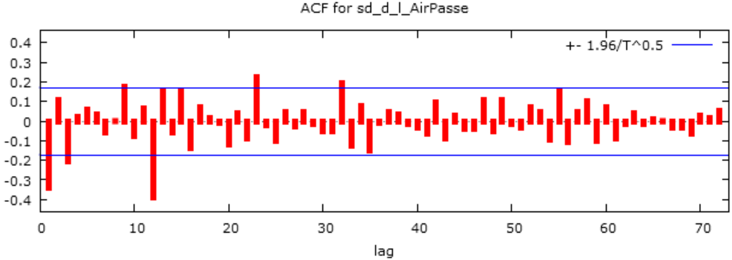
- **Ejemplo de las Aeorlíneas:** Correlograma de $(1 - L)LnY_t$

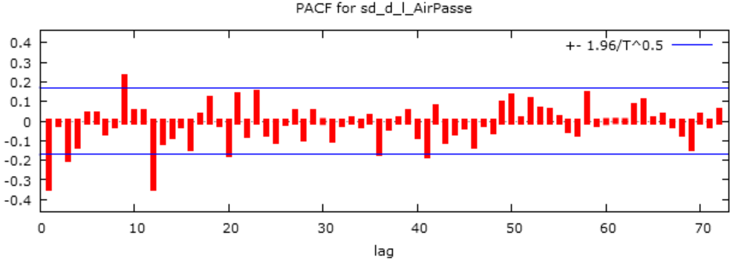




Diferencia estacional y diferencia regular (5)

- **Ejemplo de las Aeorlíneas:** Correlograma de $(1 - L)(1 - L^{12})LnY_t$





Modelo de las Aerolíneas

- Box y Jenkins (1976) utilizan estos datos y, tras identificar el modelo lo estiman como un $ARIMA(0,1,1)(0,1,1)_s$
  - A partir de entonces se conoce a ese modelo como **Modelo de las Aerolíneas**