Series de tiempo estacionarias

Fernando Borraz

Extracción de señales - dECON - 2017

Introducción

- Contamos con una serie de observaciones repetidas en el tiempo de una variable aleatoria y_t (PIB, inflación, M, etc)

- Una serie de tiempo (ST) es un conjunto de observaciones de la misma variable que son consideradas como una sucesión de variables aleatorias. $\{y_t\} = \{y_1, y_2, ..., y_T\}$ o $\{y_t\}_1^T$

- Consideramos a c/u de las y_t como variable aleatoria. Por ej. $PIB_{2016} = 1581$ mil millones de pesos es una realización entre todas las posibles

Introducción

- ¿Por qué no usar el instrumental de econometría clásica que conocemos en un modelo de ST?

- El problema es que la ST que observamos es solo UNA muestra de y_t. La colección de datos observada solo es una posible realización de un proceso estocástico

- Si por ej. tenemos N-coleccciones de $\{y_t\}$ podríamos calcular los momentos de la densidad de y_t (media, varianza, etc)

- Cantidad de parámetros a estimar: T medias, T varianzas y $\frac{T*T-T}{2}$covarianzas (si T$=$ 50 tenemos 50$+$50 $+\ \frac{50^2-50}{2} = 1325$ parámetros!!)

- Obviamente son muchos parámetros y no podemos calcularlos en base a una muestra de tamaño T (que es lo que disponemos)

Introducción

- Esto nos fuerza a usar modelos paramétricos y realizar supuestos. Así limitamos la cantidad de parámetros a estimar
- Esto es una limitante de los modelos de ST
- Así limitamos la cantidad de parámetros a estimar
- Un concepto clave será estacionariedad

Introducción

- Otra distinción relevante

  - trabajamos sobre primeros momentos (medias)
  - o sobre segundos momentos (varianzas, modelos ARCH)

- Es decir, la series podrán ser no estacionarias en media y/o en varianza

Introducción

- Modelos univariados. Una sola V.A. y que cumplen una importante restricción que es ser estacionarias. Son los modelos ARMA

- Bajo ciertas condiciones el instrumental econométrico de corte transversal (CT) puede ser usado para modelos con estas variables

- Las series no estacionarias tienen un comportamiento totalmente diferente a las estacionarias que impiden el uso de la econometría clásica y exige nuevos instrumentos

- El énfasis es en el efecto de variables no estacionarias (en media o varianza) en los modelos y como probar dicha no estacionariedad (pruebas de raíces unitarias)

Introducción

- La interacción de varias variables estacionarias las veremos en los modelos de vectores autoregresivos (VAR)

- El uso de los modelos VAR pone de manifiesto las limitaciones del enfoque de ST univariado

- Modelos con variables no estacionarias. Se enfatiza tanto la prueba como la modelización. Cointegración y modelo de vectores de corrección del error (VECM)

Modelos univariados de ST estacionarias (o metodología Box-Jenkins)

- A menudo para fines predictivos (y de corto plazo) un modelo simple que describe la conducta de una variable (o un conjunto pequeño de variables) en términos de su propio pasado brinda resultados satisfactorios

- Es importante que la variable muestre dependencia temporal (afortunadamente suele suceder en las series en economía)

- La investigación empírica muestra que modelos univariados de ST (ARMA) con pocos parámetros a menudo brindan mejores predicciones de corto plazo que modelos de ecuaciones simultáneas

Introducción

- En particular se modela la serie de la siguiente manera:
  $y_t = PS_{t-1} + \varepsilon_t$
- PS_{t-1} es la parte sistemática y es una función del pasado y conocida en t
- ε_t es un shock aleatorio en t. Es impredecible en t-1
- El objetivo es predictivo. $\hat{y}_{t+1|t}$.Es la predicción de y para el período t+1 realizada con información disponible en t
- En ST univariadas la única información muestral requerida es la referida al propio fenómeno a analizar. Esto constituye la principal ventaja y limitación de este análisis

Introducción

- Ventajas de este enfoque:

  - No require predecir las variables explicativas
  $(MCO: y_t = x_t\beta + \varepsilon_t \Rightarrow \hat{y}_{t+1|t} = x_{t+1}\overset{\wedge}{\beta}$, pero $\hat{x}_{t+1|t}?)$
  - No requiere marco teórico
  - A veces no tenemos disponible datos de las variables explicativas (variables de expectativas, etc)
  - No es problema si la frecuencia de medida de las variables es diferente (por ej. PIB trimestral, IPC mensual, TC diario)

- Desventajas:

  - No permite conocer la forma en que dicha variable y_t se relaciona con otras
  - Un modelo multivariado será más eficiente por usar mayor información

ST univariados

- El objetivo es predecir una serie usando su información pasada (por ej. inflación usando únicamente información pasada de dicha serie)

- Como no tenemos varias realizaciones de los procesos juega un rol clave el concepto de estacionariedad ($\neq$ estacionalidad)

- 2 definiciones de estacionariedad (débil, estricta)

- y_t estacionariedad débil (o de segundo orden o en covarianzas) si:

  - 1) $E(y_t) = \mu \qquad t = 1, 2, ..., \infty$
  - 2) $V(y_t) = E[(y_t - \mu)(y_t - \mu)] = \sigma^2 < \infty$ finita y no depende de t
  - 3) $E[(y_t - \mu)(y_{t-s} - \mu)] = \gamma_{t-s-t} = \gamma_s = \gamma_{-s}$ para todo t,t-s $(s \geqslant 1)$. La COV solo depende de la diferencia entre t y t-s y no directamente de t

ST univariados

- Si el proceso es estacionario débil, todas las medias y varianzas son iguales y todas la covarianzas depende únicamente de la diferencia entre t y t-s

- Las covarianzas γ_s suelen denominarse autocovarianzas

- Como las covarianzas depende de la distancia "s" (entre 2 observaciones) y no directamente de t$\Rightarrow$podemos hacer una "muestra "con los datos de solo un universo

- Series económicas estacionarias?

- Como las autocov γ_s dependen de las unidades de medida de y_t es más conveniente trabajar con las autocorrelaciones: $\tau_s = \frac{\gamma_s}{\gamma_0}, s = 0, 1, 2, ...$Si graficamos τ_s contra s obtenemos la función de autocorrelación o correlograma (ACF)

ST univariados

- y_t estacionario estricto si la función de probabilidades conjuntas de las realizaciones de y_t no depende del tiempo

- Esto es: $P\left[y_{t1} \leq b_1, y_{t2} \leq b_2, ..., y_{tn} \leq b_n\right] =$
$P\left[y_{t1+s} \leq b_1, y_{t2+s} \leq b_2, ..., y_{tn+s} \leq b_n\right]$

- Es decir, la probabilidad medida por la secuencia $\{y_t\}$ es la misma que aquella medida por $\{y_{t+s}\}$

- La probabilidad que y pertenezca a cierto intervalo es la misma independientemente en que momento (pasado o futuro) la calculamos

- Ej. la probabilidad que la inflación sea menor a cierto valor por ej 5% es la misma entre y_{2016}, y_{2015} que entre y_{1901}, y_{1900}

Estacionariedad

- Notar que estacionariedad estricta $\nRightarrow$estacionariedad débil pues los segundos momentos pueden no existir

- Ej. distribución Cauchy, $f(y) = \frac{1}{\pi(1+y^2)}$, varianza de y_t es ∞

- Estacionariedad estricta + segundos momentos finitos$\Rightarrow$ estacionariedad débil

- Si $y_t \sim$normal, estacionariedad estricta $\Leftrightarrow$ estacionariedad débil

- En el curso asumiremos el supuesto de normalidad por lo que simplemente nos referiremos a estacionariedad (que en este caso es débil o estricta)

- ¿Cómo sabemos si la serie y_t es estacionaria? Es un tema fundamental. Análisis gráfico, ACF y pruebas de raíces unitarias

Ergodicidad

- Una serie y_t estacionaria es ergódica en la media si:
  $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_t = E\left(y_t\right)$
- Esto es, la media muestral converge a la media del proceso cuando la muestra crece
- En la práctica se requiere que las autocov tiendan a 0.
  $\left(\sum_{j=0}^{\infty} \mid \gamma_j \mid < \infty\right)$
- Intuición. Cuando $s \to \infty$ el pasado remoto no debe contener información útil para las predicciones actuales del proceso
- Ergodicidad y estacionariedad no siempre son equivalentes
- Un proceso puede ser estacionario pero no ergódico

Proceso de ruido blanco

- Un proceso ruido blanco se caracteriza por no tener estructura dicernible

- y_t ruido blanco si:
  - 1) $E\left(y_t\right) = 0 \qquad t = 1, 2, ..., \infty$
  - 2) $V\left(y_t\right) = \sigma^2 < \infty$ finita y no depende de t
  - 3) $E\left[\left(y_t - \mu\right)\left(y_{t-s} - \mu\right)\right] = 0 \qquad \left(s \geqslant 1\right)$

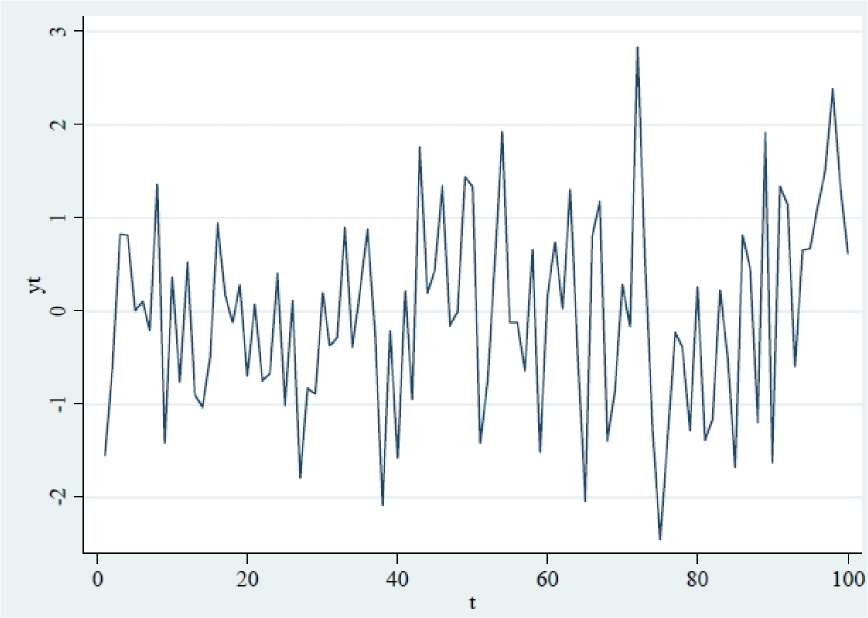
- Condicional en la información disponible en t-1 no es posible predecir el valor futuro del proceso $\left(\hat{y}_{t+1|t} = 0\right)$

- ¿Un proceso ruido blanco es estacionario?

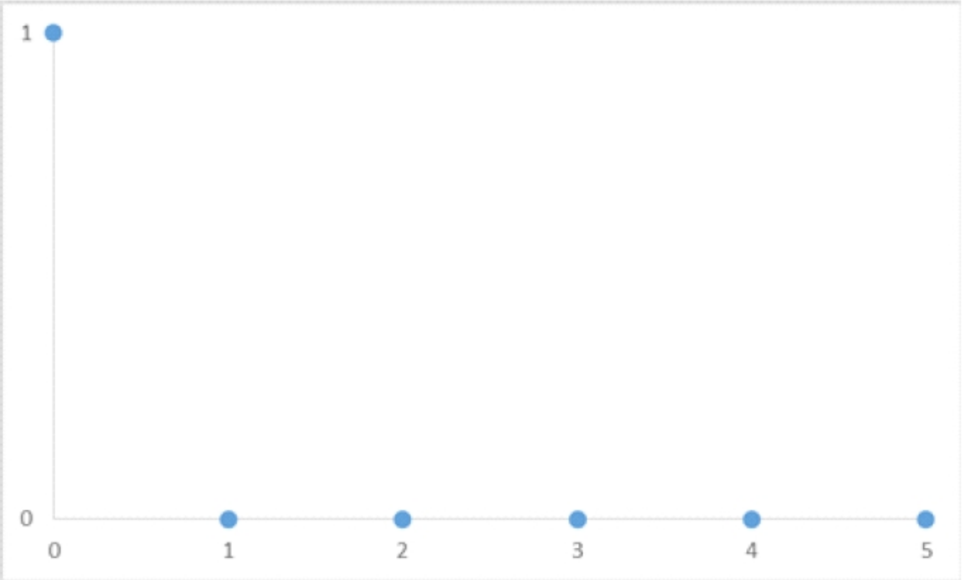
- Recordar que requeríamos que el error del modelo fuera un proceso de ruido blanco

- Por ejemplo, en $\boxed{y_t = PS_{t-1} + \varepsilon_t}$,ε_t debe ser un proceso ruido blanco

Proceso de ruido blanco: ejemplo



Proceso de ruido blanco: ACF



Conceptos básicos

- Gran parte del análisis univariado de DT lo realizaremos usando la función de autocorrelación del proceso

- $\tau_s = \frac{\gamma_s}{\gamma_0} \qquad s = 0, 1, 2, ...$

- La función de autocorr identifica y caracteriza un proceso estocástico completamente

- Esto es, si dos procesos tienen la misma función de autocorr $\Rightarrow$son el mismo proceso

- Para el análisis empírico nos basaremos en las autocorr

Análisis empírico univariado estacionario

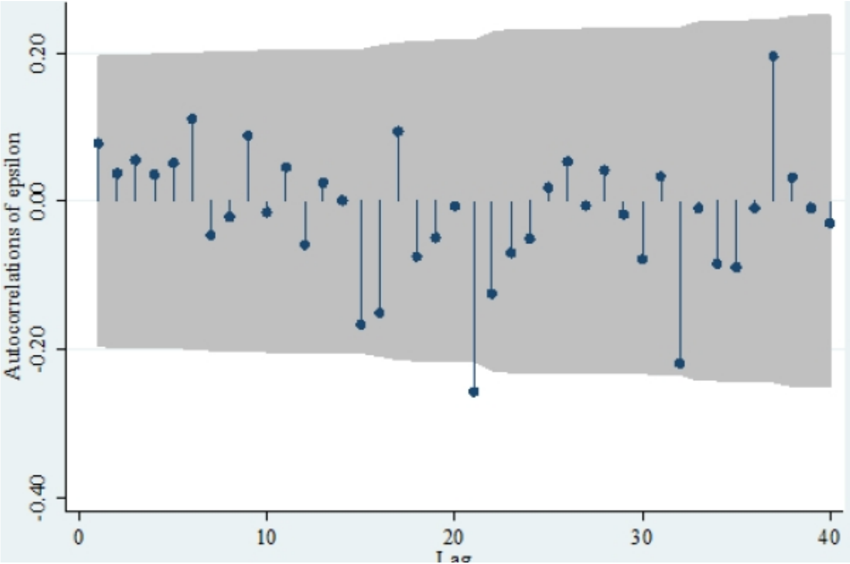
- En la práctica no conocemos media, var y autocorr y debemos estimarlos en base a una muestra de tamaño T

- Es decir, estimamos la función de autocorrelación

- $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$
- $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(y_t - \bar{y} \right)^2$
- $\hat{\tau}_s = \frac{\sum_{t=s+1}^{T} \left(y_t - \bar{y} \right) \left(y_{t-s} - \bar{y} \right)}{\sum_{t=1}^{T} \left(y_t - \bar{y} \right)^2}$

- Inferencia? Precisamos pruebas de hipótesis para estos estadísticos estimados

ACF empírica: ruido blanco?



Análisis empírico univariado estacionario

- $\overset{\wedge}{\tau}_s \overset{a}{\sim} N\left(0, \frac{1}{T}\right)$
- Un intervalo de confianza al $(1-\alpha)\%$ para $\overset{\wedge}{\tau}_s$viene dado por $\overset{\wedge}{\tau}_s \pm z_{1-\frac{\alpha}{2}} * \frac{1}{\sqrt{T}}$
- Donde α es el nivel de significancia de la prueba y $z_{\left(1-\frac{\alpha}{2}\right)}$ es el valor de tablas de una normal estandar que acumula $1-\frac{\alpha}{2}$
- Si el coeficente de correlación muestral para un determinado valor de s cae fuera del intervalo entonces rechazamos la hipótesis nula que la correlación en el lag s es 0

Análisis empírico univariado estacionario

- También es posible hacer la prueba de hipótesis que s de los coeficientes de correlación son simultáneamente 0 usando el estadístico Q de Box y Pierce (1970)

- $H_0) \tau_1 = \tau_2 = ... = \tau_s = 0$
- $Q = T \sum_{i=1}^{s} \overset{\wedge}{\tau}_i^2 \overset{a}{\sim} \chi_s^2$

- Se eleva al cuadrado para que los coeficientes de correlación positivos y negativos no se cancelen

- Q es una suma de normales al cuadrado

- Q no funciona adecuadamente en muestras pequeñas

Análisis empírico univariado estacionario

- Ljung y Box (1978) modifican la prueba Q con mejores propiedades en muestras pequeñas

- $Q^* = T\,(T+2)\sum\limits_{i=1}^{s} \frac{\overset{\wedge}{\tau}{}_i^2}{T-i} \overset{a}{\sim} \chi_s^2$

- Para T grande (T$\rightarrow \infty$): T+2 se cancela con T-i y $Q \cong Q^*$

- Para que una serie sea ruido blanco no deberíamos rechazar H_0)

Modelos univariados para procesos estacionarios

- De acuerdo a la información utilizada por definir los siguientes modelos:
- i) $y_t = f(innovaciones\ pasadas) + \varepsilon_t \Rightarrow$Modelos Moving Average (MA)
- ii) $y_t = f(valores\ pasados\ de\ la\ serie)\ + \varepsilon_t \Rightarrow$Modelos Autoregressive (AR)
- iii) $y_t = f(innovaciones\ pasadas, valores\ pasados) + \varepsilon_t \Rightarrow$Modelos ARMA

Procesos de media movil (MA)

- Se construyen usando una combinación lineal de ruidos blancos

- Un proceso de medias móviles de orden finito q, MA(q) se define como: $y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$

- Para ver sus propiedades (media, var, autocorr) comencemos por el MA(1): $y_t = \alpha + \varepsilon_t + \theta \varepsilon_{t-1}$

- $E(y_t) = E(\alpha + \varepsilon_t + \theta \varepsilon_{t-1}) = \alpha + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) = \alpha$

- $V(y_t) = E\left[(y_t - E(y_t))^2\right] = E\left[(y_t - \alpha)^2\right] =$ $E\left[(\alpha + \varepsilon_t + \theta \varepsilon_{t-1} - \alpha)^2\right] = E\left[(\varepsilon_t + \theta \varepsilon_{t-1})^2\right]$

- $= E\left(\varepsilon_t^2\right) + 2\theta_1 E(\varepsilon_t \varepsilon_{t-1}) + \theta^2 E\left(\varepsilon_{t-1}^2\right) = \sigma^2\left(1 + \theta^2\right) = V(y_t) = \gamma_0$

MA(q)

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

- $E\left(y_t\right) = \alpha$

- $V\left(y_t\right) = \sigma^2\left(1 + \theta_1^2 + \theta_2^2 + .. + \theta_q^2\right)$

- $\gamma_s = E\big[\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + .. + \theta_q \varepsilon_{t-q}\right)$
  $\left(\varepsilon_{t-s} + \theta_1 \varepsilon_{t-s-1} + \theta_2 \varepsilon_{t-s-2} + .. + \theta_q \varepsilon_{t-s-q}\right)\big]$

- $= E\left(\theta_s \varepsilon_{t-s}^2 + \theta_{s+1}\theta_1 \varepsilon_{t-s-1}^2 + \theta_{s+2}\theta_1 \varepsilon_{t-s-2}^2 + .. + \theta_q \theta_{q-s} \varepsilon_{t-q}^2\right) =$
  $\sigma^2\left(\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_1 + .. + \theta_q \theta_{q-s}\right) \qquad s = 1, 2, .., q$

MA(q)

- $\tau_0 = 1$
- $\tau_s = \frac{\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_1 + .. + \theta_q\theta_{q-s}}{1+\theta_1^2+\theta_2^2+..+\theta_q^2} \qquad s = 1, 2, .., q$
- $\tau_s = 0 \qquad s > q$
- La acf de un MA(q) muere a partir de q
- La información del pasado mayor a q no tiene efecto en la predicción de y_t
- MA(q) siempre es estacionario
- Construimos IC para ver si los $\overset{\wedge}{\tau}_s$ son estadísticamente diferentes de 0

MA infinito

- ¿Qué pasa si q$\rightarrow \infty$?
- $y_t = \alpha + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + .. = \alpha + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \qquad \psi_0 = 1$
- Estacionariedad?

MA infinito

- *Teorema*: Si la secuencia de coeficientes del MA(∞) es square summable $\left(\sum_{i=0}^{\infty} \psi_i^2 < \infty \right)$, entonces la secuencia de autocov del MA(∞) es sumable y el proceso es estacionario

- $V\left(y_t\right) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$

- En la práctica requerimos una condición más estricta pero más conveniente de probar que es absolutely summability $\sum_{i=0}^{\infty} \mid \psi_i \mid < \infty$

Procesos autoregresivos (AR)

- Los procesos AR tienen una lógica diferente pues no se construyen a partir de la historia de los shocks sino a partir de los valores pasados de la serie (más un shock estocástico contemporáneo)

- AR(p): $y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + .. + \phi_p y_{t-p} + \varepsilon_t$

- Tomemos el caso más simple que es el AR(1): $y_t = \alpha + \phi y_{t-1} + \varepsilon_t$

- Estacionariedad? Veamos sus propiedades

- Para calcular los momentos debemos describir a y_t como una función de los ε pues conocemos su E, VAR, Cov

- $y_t = \alpha + \phi y_{t-1} + \varepsilon_t \Rightarrow y_{t-1} = \alpha + \phi y_{t-2} + \varepsilon_{t-1} \Rightarrow y_t = \alpha + \phi\left(\alpha + \phi y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t$

- Despejamos y_{t-2}, y_{t-3},etc

AR(1)

- $y_t = \alpha + \phi\left(\alpha + \phi y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t = \alpha + \alpha\phi + \phi^2 y_{t-1} + \phi\varepsilon_{t-1} + \varepsilon_t$
- $= \alpha\left(1 + \phi\right) + \phi^2\left(\alpha + \phi y_{t-3} + \varepsilon_{t-2}\right) + \phi\varepsilon_{t-1} + \varepsilon_t$
- $= \alpha\left(1 + \phi + \phi^2\right) + \phi^3 y_{t-3} + \phi^2\varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \Rightarrow$
- $\boxed{y_t = \alpha\sum_{i=0}^{t-1}\phi^i + \phi^t y_0 + \sum_{i=0}^{t-1}\phi^i\varepsilon_{t-i}}$
- Estacionariedad de y_t?

AR(1)

- $y_t = \alpha \sum_{i=0}^{t-1} \phi^i + \phi^t y_0 + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}$
- Por lo que nos queda un MA(∞) (representación dual de modelos dinámicos)
- Vimos que para que el proceso sea estacionario la secuencia tienen que ser absolute summable. En este caso se requiere que $\mid \phi \mid < 1$
- Notar que si la muesta no es infinita o lo suficientemente grande el proceso dependerá del "principio de los tiempos", es decir de y_0.Por lo que hay que tener cuidado cuando el tamaño muestral es pequeño
- Si la muestra es grande y $\mid \phi \mid < 1$ (o asumimos $y_0 = 0$)$\Rightarrow \phi^t y_0 \rightarrow 0$
- Además, si la muestra es grande y $\mid \phi \mid < 1 \;\Rightarrow \alpha \sum_{i=0}^{t-1} \phi^i = \alpha \left(1 + \phi + \phi^2\right) \rightarrow \frac{\alpha}{1-\phi}$

AR(1)

- $y_t = \frac{\alpha}{1-\phi} + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}$

- $\mid \phi \mid > 1$ no tiene sentido económico pues implicaría que las innovaciones pasadas son más importantes que las cercanas para explicar y_t (y_t explosiva). La media y la varianza crecen con t

- La estacionariedad de la serie depende del coeficiente ϕ

- ($\phi = 1$, Random walk):
  $y_t = \alpha + y_{t-1} + \varepsilon_t \Rightarrow y_t = \alpha + (\alpha + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = ...$

$= \sum_{i=0}^{t-1} \alpha + y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \Rightarrow y_t = \alpha t + y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}$ Crece con el paso del tiempo. Shocks permanentes

Propiedades AR(1)

- $\gamma_1 = \frac{\sigma^2 \phi}{1-\phi^2} \qquad \Rightarrow \tau_1 = \frac{\gamma_1}{\gamma_0} = \left(\frac{\sigma^2 \phi}{1-\phi^2}\right) / \left(\frac{\sigma^2}{1-\phi^2}\right) = \phi$

- Idem para $\gamma_2, \tau_2 . \tau_2 = \phi^2 \qquad \tau_s = \phi^s$

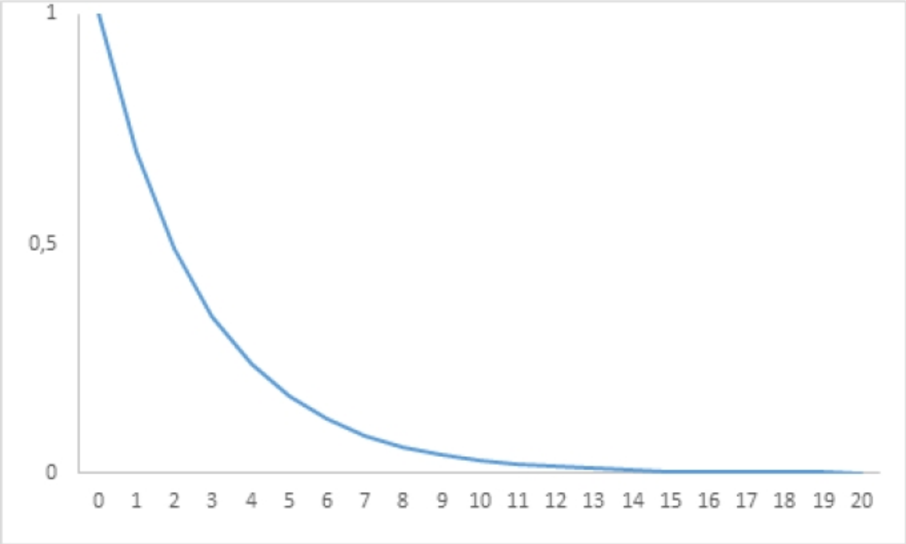
- Por lo que la acf decae exponencialmente (con cambio de signo si $\phi < 0$)

- Si el proceso es estacionario más lento o más rápido pero los shocks se disipan

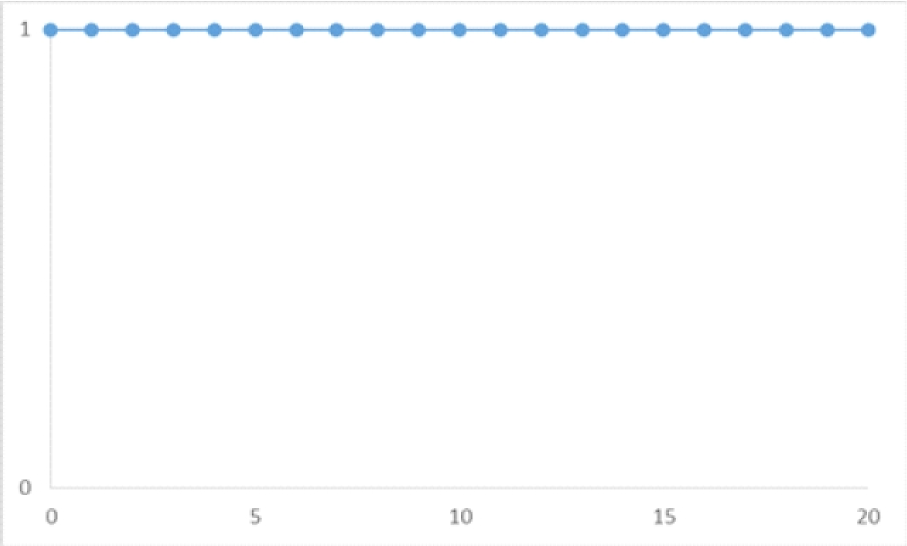
- Si el proceso no es estacionario los shocks no se disipan

- Sin embargo en una muestra pequeña es difícil distinguirlos pues puede que se confunda persistencia con lenta disipasión

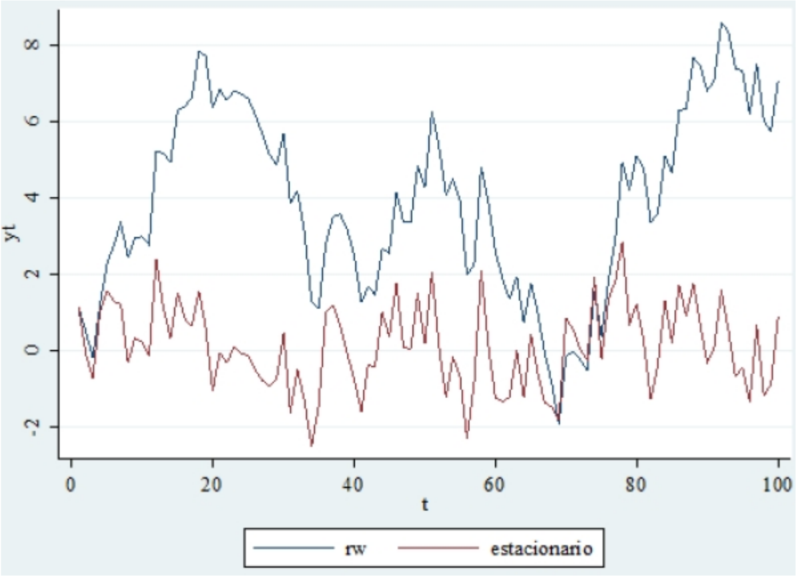
AR(1) estacionario coeficiente positivo: ACF



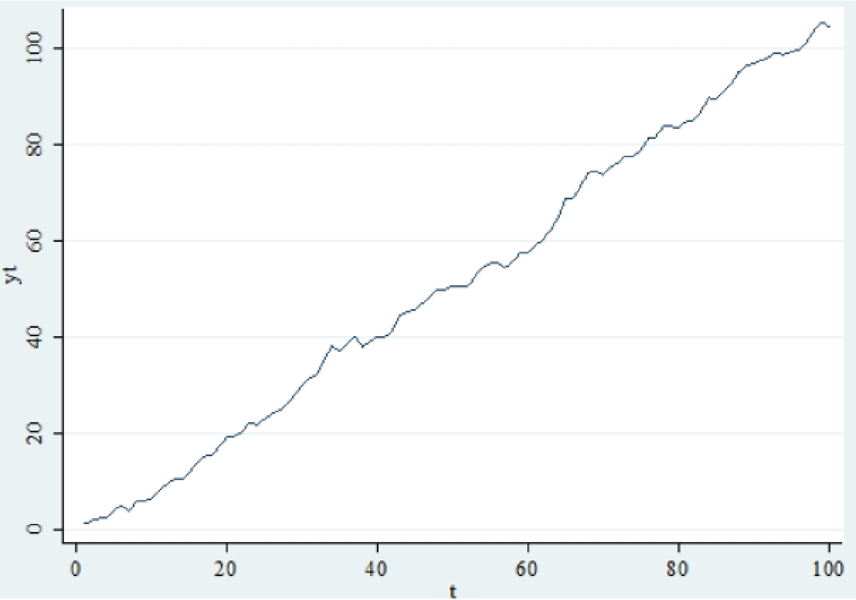
AR no estacionario (RW): acf



AR(1) estacionario y no estacionario (random walk)



RW con constante



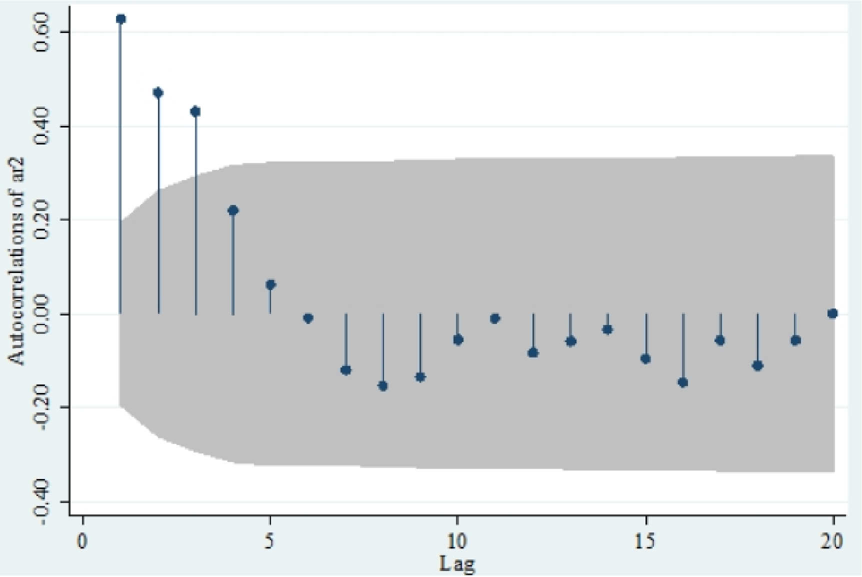
AR(p)

- De igual manera podemos calcular la acf para un AR(2), AR(3), ..,AP(p)
- Sin embargo es engorroso su calculo
- Una manera más sencilla de obtener las acf es mediante las ecuaciones de Yule-Walker

AR(2): acf

- Mismo patrón que el acf del AR(1): decae geometricamente
- En un AR(2) y_{t-3} esta correlacionado con y_t pues afecta a $y_{t-2.}$ Idém para $y_{t-4}, y_{t-5}, ..$
- Idém para AR(3),..,AR(p)
- Por lo que mediante el análisis gráfico de la acf no podemos determinar el orden p del proceso AR
- Para determinar p utilizamos una función adicional: función de autocorrelación parcial (pacf)

AR(2): acf



0.60
0.40
0.20
0.00
-0.20
-0.40
Autocorrelations of ar2
0
5
10
15
20
Lag

Función de autocorrelación parcial (pacf)

- Para identificar el orden p del AR usaremos dos correlogramas: total (acf) y parcial (pacf)

- La pacf nos dice cual es la correlación entre y_t, y_{t-s}menos aquella parte que se puede explicar por una proyección lineal entre $y_t, y_{t-(s-1)}$

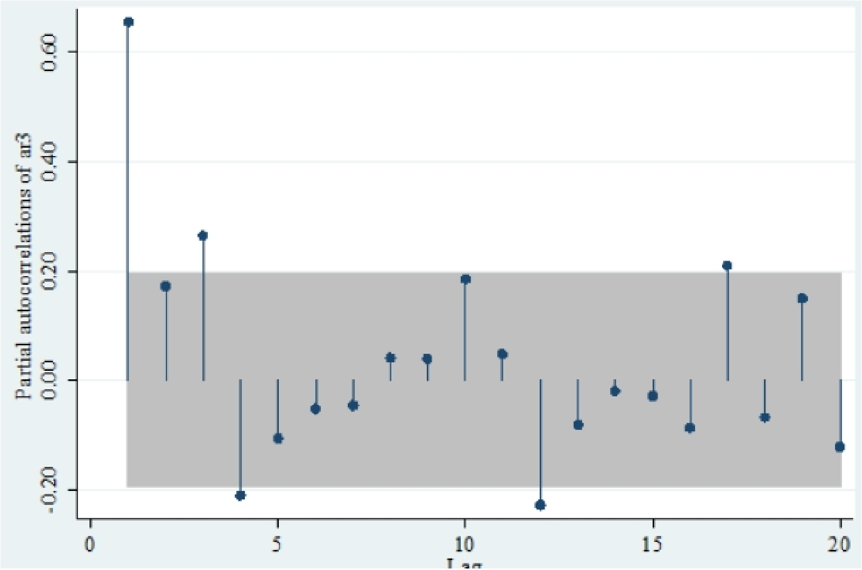
- $\tau_s^* = Corr\left[y_t - E\left(y_t \mid y_{t-1}, y_{t-2}, .., y_{t-(s-1)}\right); y_{t-s}\right]$

- La correlación parcial entre y_t, y_{t-s} es la correlación simple entre y_t, y_{t-s} su predicción realizada con s-1 rezagos

- Por lo que para un AR(p), la pacf debería sería 0 a partir de $p+1$

- Para el primer rezago acf=pacf pues no hay efectos de rezagos intermedios que eliminar: $\tau_1 = \tau_1^*$

AR(3): PACF empírica



Estacionariedad del AR(p)

- Para $p = 1$ el proceso es estacionario si $\mid \phi \mid < 1$
- El proceso AR(p) es estacionario?
- Para responder este usaremos el operador de rezagos L que permite operar con los rezagos de manera más cómoda
- $Ly_t = y_{t-1}$ $\quad L^2 y_t = L\left(Ly_t\right) = Ly_{t-1} = y_{t-2}$ $\quad L^s y_t = y_{t-s}$
- $\Delta y_t = y_t - y_{t-1} = \left(1 - L\right) y_t$ $\quad \Delta^s y_t = y_t - y_{t-s} = \left(1 - L^s\right) y_t$
- El operador de rezagos funciona de igual forma que el operador "multiplicar por"

Operador de rezagos

- $L\alpha = \alpha \qquad L\alpha y_t = \alpha L y_t = \alpha y_{t-1}$
- $\left(L^i + L^j\right) y_t = L^i y_t + L^j y_t = y_{t-i} + y_{t-j}$
- $L^i L^j y_t = L^i \left(L^j y_t\right) = L^i y_{t-j} = y_{t-j-i}$
- $L^{-i} y_t = y_{t+i}$
- AR(1): $y_t = \alpha + \phi y_{t-1} + \varepsilon_t : (1 - \phi L)\, y_t = \alpha + \varepsilon_t$

Operador de rezagos

- $(1 - \phi L)\, y_t = \alpha + \varepsilon_t$

- $y_t = \frac{\alpha}{(1-\phi L)} + \frac{\varepsilon_t}{(1-\phi L)}$
- Si $\mid \phi \mid < 1 \Rightarrow y_t = \left(1 + \phi L + \phi^2 L^2 + ..\right) \alpha + \left(1 + \phi L + \phi^2 L^2 + ..\right) \varepsilon_t$
- Esto ya lo vimos. Pasamos de un AR(1) estacionario a un MA(∞)

Estacionariedad del AR(p)

- $y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + .. + \phi_p y_{t-p} + \varepsilon_t$

- $\left(1 - \phi_1 L - \phi_2 L^2 - .. - \phi_p L^p\right) y_t = \alpha + \varepsilon_t$

- α, ε_t son estacionarios por lo que no afectan la tendencia de la serie y por lo tanto no afectan su estacionariedad

- Por lo que para analizar la estacionariedad solo importa la parte homógenea:

- $\left(1 - \phi_1 L - \phi_2 L^2 - .. - \phi_p L^p\right) y_t = 0$

- Ecuación en diferencias de orden p

Estacionariedad del AR(p)

- $\left(1-\phi_1 L-\phi_2 L^2-..-\phi_p L^p\right) y_t=0$
- Guess: $y_t=A\lambda^t \Rightarrow$
- $\left(1-\phi_1 L-\phi_2 L^2-..-\phi_p L^p\right) A\lambda^t=0$
- $A\lambda^t-\phi_1 A\lambda^{t-1}-\phi_2 A\lambda^{t-2}-..\phi_p A\lambda^{t-p}=0$ Dividiendo por $A\lambda^{t-p}$
- $\frac{A\lambda^t}{A\lambda^{t-p}}-\frac{\phi_1 A\lambda^{t-1}}{A\lambda^{t-p}}-\frac{\phi_2 A\lambda^{t-2}}{A\lambda^{t-p}}-..-\frac{\phi_p A\lambda^{t-p}}{A\lambda^{t-p}}=0$
- $\lambda^p-\phi_1\lambda^{p-1}-\phi_2\lambda^{p-2}-..-\phi_p\lambda^0=0$
- p raíces características de la ecuación característica: $\lambda_1, \lambda_2, .., \lambda_p$

Estacionariedad del AR(p)

- Notar que las soluciones no son únicas. Cualquier combinación lineal también es solución
- La solución de la ecuación homógenea es
- $y_t = A_1\lambda_1^t + A_2\lambda_2^t + .. + A_p\lambda_p^t$
- La solución concreta va a depender de las constantes A_i pero no la estacionariedad de la serie
- Cuando $t \to \infty$ la trayectoria viene dada por la mayor raíz en valor absoluto
- Si existe al menos una raíz $\mid \lambda_i \mid > 1 \Rightarrow y_t$ no tiene límite finito y por lo tanto no es estacionario

Estacionariedad del AR(p)

- Por lo que el AR(p) es estacionario si todas las raíces de la ecuación característica son menores a 1 en valor absoluto
- Si hay soluciones complejas lo mismo vale para el modulo $(\lambda = a + bi,\ \sqrt{a^2 + b^2} < 1,\ i = \sqrt{-1})$
- Ejemplo: $y_t = 4 + 1,1y_{t-1} - 0,18y_{t-2} + \varepsilon_t$
- Raíces polinomio AR: 5 y 1,1$\Rightarrow$estacionario

Procesos ARMA(p,q)

- Nada determina que no podamos definir un proceso con componentes AR y MA. Son los procesos ARMA(p,q)
- $y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + .. + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + .. + \theta_q \varepsilon_{t-q} + \varepsilon_t$
- Calcular la acf es tedioso por la gran cantidad de parámetros
- La estacionariedad del ARMA (p,q) depende de la estacionariedad del componente AR(p)
- Usaremos los acf y pacf para determinar el orden p,q del proceso

Teorema de representación de Wold

- ¿Qué nos permite suponer que es posible representar series económicas mediante modelos ARMA?

- Aunque tengamos una secuencia de valores que sean consistentes con una acf

- *Teorema de Wold*: Cualquier proceso debilmente estacionaro (o de segundo orden) con media 0 puede ser representado de modo único por un componente determístico lineal y un componente no determínistico lineal

- $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + .. + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \qquad \left(\psi_0 = 1; \sum_{i=0}^{\infty} \psi_i^2 < \infty\right)$

- Esto nos permite racionalizar los modelos ARMA

- Notar que no se requiere $\varepsilon_t \sim N$ o $\varepsilon_t \sim iid$. La representación de la serie es única entre modelos lineales

ARMA o metodología Box-Jenkins (1976)

- Apareció en los 70 como una alternativa ante el fracaso predictivo de los modelos econométricos de gran escala como los de ecuaciones simultáneas. 6 pasos
- 1. Transformar las series en estacionarias (sino lo son) pues dicha metodología se aplica únicamente a a series estacionarias
- 2. Eliminar ciertos componentes determínisticos como la estacionalidad y valores atípicos
- 3. Identificar el modelo. Determinar p,q
- 4. Estimar los parámetros del modelos $(\phi's, \theta's, \alpha, \sigma^2)$
- 5. Realizar pruebas de especificación. Errores ruido blanco
- 6. Realizar predicciones con base al modelo estimado

ARMA o metodología Box-Jenkins

- Estimación de los parámetros $\theta = \left(\alpha, \phi_1, \phi_2, .., \phi_p, \theta_1, \theta_2, .., \theta_q, \sigma^2\right)'$. No es posible usar MCO. Estimación por MV

- $y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + .. + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + .. + \theta_q \varepsilon_{t-q} + \varepsilon_t$

- Asumimos $\varepsilon_t \sim N\left(0, \sigma^2\right)$

- $\varepsilon_t = y - \alpha - \phi_1 y_{t-1} - .. - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - .. - \theta_q \varepsilon_{t-q}$

- $L(\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{\varepsilon_t^2}{2\sigma^2}} = \frac{1}{(2\Pi)^{\frac{T}{2}} (\sigma^2)^{\frac{T}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2}$
- $LnL(\theta) = -\frac{T}{2} Ln(2\Pi) - \frac{T}{2} Ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2$

- Max $LnL(\theta)$ con respecto al vector de parámetros θ

ARMA o metodología Box-Jenkins: estimación

- $LnL(\theta) = -\frac{T}{2} Ln\,(2\Pi) - \frac{T}{2} Ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2$

- $\varepsilon_t = y_t - \alpha - \phi_1 y_{t-1} - .. - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - .. - \theta_q \varepsilon_{t-q}$

- Notar:

- i) No observamos $y_0, y_{-1}, .., y_{-p+1}, \varepsilon_0, \varepsilon_{-1}, .., \varepsilon_{-q+1}$

- ii) El problema de maximización no tiene una solución cerrada. No despejable analíticamente

ARMA o metodología Box-Jenkins: estimación

- i) Asumir que los $\varepsilon's$ no observados toman su valor medio
- Esto es: $\varepsilon_0 = \varepsilon_{-1} = .. = \varepsilon_{-q+1} = 0$
- Con respecto a los y no observados se puede asumir que toman su valor esperado pero Box y Jenkins recomiendan usar los valores que efectivamente tomaron y por lo tanto inician la suma en $t = p + 1$ y no en $t = 1$
- Por lo que solo se precisa asumir $\varepsilon_p = \varepsilon_{p-1} = .. = \varepsilon_{p-q+1} = 0$
- $LnL(\theta) = -\frac{(T-p)}{2} Ln\left(2\Pi\right) - \frac{(T-p)}{2} pLn\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=p+1}^{T} \varepsilon_t^2$

ARMA o metodología Box-Jenkins: estimación

- ii) $LnL(\theta) = -\frac{(T-p)}{2} Ln\left(2\Pi\right) - \frac{(T-p)}{2} pLn\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=p+1}^{T} \varepsilon_t^2$

- La maximización no tiene solución cerrada

- Solución numérica mediante el uso de algoritmos. Procedimiento iterativo

- Por ejemplo Newton Raphson

- $\theta_{j+1} = \theta_j - G_j^{-1} g\left(\theta_j\right)$ $\qquad$ $j =$iteración

- g es el vector de derivadas (gradiente) de la función a maximizar

- G es el hessiano de la función a maximizar

ARMA o metodología Box-Jenkins: estimación mediante algoritmos

- Limitaciones:

  - i) Valores iniciales? aleatorios?

  - ii) Puede diverger con funciones complejas y lejos del óptimo

  - iii) Puede diverger si G^{-1} no es definido negativo

  - iv) Convergencia a máximo local y no global

ARMA o metodología Box-Jenkins: pruebas de especificación

- Criterio de información de Akaike (AIC):

$$Min_m TLn\,(SCR) + 2m \qquad m = p + q + 1$$

- Criterio de información bayesiano de Schwartz (SBC):

$$Min_m TLn\,(SCR) + Ln(T)m$$

- Elegimos p,q tales que los criterios de información tomen el valor más pequeño

- Notar que como $Ln\,(T) > 2 \Rightarrow$SBC selecciona modelos más parsimoniosos que AIC

- Los resultados no tienen porque ser coincidentes según ambos criterios

ARMA o metodología Box-Jenkins: pruebas de especificación

- SBC tiene mejores propiedades que AIC en muestras grandes. Es asintóticamente consitente pero ineficiente
- AIC esta sesgado a elegir un modelo sobreparametrizado pero es más eficiente que SBC
- Notar que los criterios de información nos dan otra herramienta para seleccionar el orden p,q en el paso 3
- Existen otros criterios que varían en la ponderación que le dan a cada término

ARMA o metodología Box-Jenkins: predicción

- Es la finalidad de los modelos ARMA

- También es una prueba importante de lo adecuado que es el modelo estimado

- Dos tipos de predicciones:
  - i) Dentro de la muestra. Nos sirven para evaluar el modelo
  - ii) Fuera de la muestras. Son las predicciones interesantes

- Por ej. si tenemos datos mensuales de inflación de 1980M1 a 2017M5 podemos usar dicha información para construir el modelo o dejar algunas observaciones fuera

- Estimar el modelo para el período 1980M1 a 2016M12

- Usar las observaciones del período 2017M1 a 2017M5 para evaluar la calidad predictiva del modelo

ARMA o metodología Box-Jenkins: predicción

- El modelo usado para la predicción es
- $f_{t,s} = \alpha + \sum_{i=1}^{p} \phi_i f_{t,s-i} + \sum_{j=1}^{q} \theta_j f_{t,s-j} \varepsilon_{t-j}$ donde
- $f_{t,s} = y_{t+s} \qquad s \leq 0$
- $\varepsilon_{t,s} = 0 \qquad s > 0$
- $\varepsilon_{t,s} = \varepsilon_{t,s} \qquad s \leq 0$

Predicción con modelos MA(q)

- Un MA(q) solo tiene memoria q
- Por ej. para un MA(3): $y_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t$
- Estamos en el período t y queremos predecir 1,2,..,s períodos adelante
- Conocemos $y_t, y_{t-1}, .., \varepsilon_t, \varepsilon_{t-1}, ...$

Prediciendo con un AR(1) (estacionario)

- $y_t = \alpha + \phi y_{t-1} + \varepsilon_t \Rightarrow y_{t+1} = \alpha + \phi y_t + \varepsilon_{t+1}$
- $E_t\left(y_{t+1}\right) = y_{t+1|t} = \alpha + \phi y_t$
- $y_{t+2|t} = \alpha + \phi y_{t+1|t} = \alpha + \phi\left(\alpha + \phi y_t\right) = \alpha\left(1+\phi\right) + \phi^2 y_t$
- ....
- $y_{t+s|t} = \alpha\left(1 + \phi + \phi^2 + .. + \phi^{s-1}\right) + \phi^s y_t$
- $\lim_{s\to\infty} y_{t+s|t} = \frac{\alpha}{1-\phi}$

Prediciendo con un AR(1)

- $\varepsilon_{t+1|t} = y_{t+1} - y_{t+1|t} = \alpha + \phi y_t + \varepsilon_{t+1} - (\alpha + \phi y_t) = \varepsilon_{t+1}$
- $V\left(\varepsilon_{t+1|t}\right) = \sigma^2$
- $\varepsilon_{t+2|t} = y_{t+2} - y_{t+2|t} = \alpha + \phi y_{t+1} + \varepsilon_{t+2} - \left[\alpha\left(1+\phi\right) + \phi^2 y_t\right]$
- $= \alpha + \phi\left(\alpha + \phi y_t + \varepsilon_{t+1}\right) + \varepsilon_{t+2} - \left[\alpha\left(1+\phi\right) + \phi^2 y_t\right]$
- $= \phi\varepsilon_{t+1} + \varepsilon_{t+2}$
- $V\left(\varepsilon_{t+2|t}\right) = \sigma^2\left(1+\phi^2\right)$

Prediciendo con un AR(1)

- ....

- $\varepsilon_{t+s|t} = y_{t+s} - y_{t+s|t} = \varepsilon_{t+s} + \phi\varepsilon_{t+s-1} + \phi^2\varepsilon_{t+s-2} + .. + \phi^{s-1}\varepsilon_{t+1}$

- $V\left(\varepsilon_{t+s|t}\right) = \sigma^2\left(1 + \phi^2 + \phi^4 + .. + \phi^{2(s-1)}\right)$

- La varianza del error de predicción crece con el horizonte de predicción pero a una tasa decreciente

- $\lim_{s\to\infty} V\left(\varepsilon_{t+s|t}\right) = \frac{\sigma^2}{1-\phi^2}$

- No solo nos interesa la predicción puntual sino un intervalo de confianza (IC) asociado a la misma

Prediciendo con un AR(1)

- La predicción a un paso es $\alpha + \phi y_t$
- Si $\varepsilon_t \sim N\left(0, \sigma^2\right)$ un IC la 95% para dicha predicción es:
- $(\alpha + \phi y_t) \pm 1,96\sigma$
- Esto es: $[\alpha + \phi y_t - 1,96\sigma; \alpha + \phi y_t + 1,96\sigma]$
- Para $s = 2$ la predicción es
  $\alpha\,(1+\phi) + \phi^2 y_t \qquad V(\varepsilon_{t+2|t}) = \sigma^2\left(1+\phi^2\right)$
- IC: $\alpha\,(1+\phi) + \phi^2 y_t \pm 1,96\sigma\sqrt{1+\phi^2}$

Evaluación de la predicción

- Min MSE (Mean Square Error)$= \frac{1}{s}\sum_{i=1}^{s}\left(y_{t+i} - y_{t+i|t}\right)^2$
- Min MAE (Mean Absolute Error)$= \frac{1}{s}\sum_{i=1}^{s} \mid y_{t+i} - y_{t+i|t} \mid$
- Min MAPE (Mean Absolute Percent Error)$= 100 * \frac{1}{s}\sum_{i=1}^{s} \mid \frac{y_{t+i} - y_{t+i|t}}{y_{t+i}} \mid$
- Nos quedamos con el modelo que minimize cualquiera de estos criterios

Predicción

- Una manera más general de mirar el tema de predicción es a través de la función impulso respuesta (IRF)
- Damos un shock al proceso y vemos como se disipa
- $\varepsilon_t = 1$ (o un desvío estandar) $\varepsilon_{t+s} = 0$ $s > 0$
- AR(1): $y_t = \phi y_{t-1} + \varepsilon_t \Rightarrow y_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$
- Por lo que primero estimamos el parámetro ϕ y luego caculamos la respuesta del cambio en 1 en ε_t en los valores de $y_{t+1}, y_{t+2}, \ldots$
- El objetivo es determinar en cuantos períodos se disipa el shock
- Análisis gráfico de $\frac{\partial y_{t+s}}{\partial \varepsilon_t}$ para s=1,2,...

Predicción: comentarios finales

- Series sujetas a cambios estructurales o de régimen no pueden ser predecidas
- La precisión de las estimaciones cae con el horizonte de predicción
- La predicción en base al modelo econométrico no sustituye el juicio analítico
- Utilizar un modelo econométrico para predecir basado en sólidos fundamentos teóricos y complementarlo con juico e interpretación de expertos