ECONOMETRÍA II:

ECONOMETRÍA DE SERIES TEMPORALES

Modelación con ARMA

Método Box-Jenkins:

• Un libro que ha tenido una gran influencia es el de Box y Jenkins (1976): *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden-Day

• En este libro se propone un método de modelación con modelos ARMA que ha sido y sigue siendo muy utilizado

• Este método se conoce como el *método Box-Jenkins* ("Box-Jenkins methodology")

• *Nota*: Casi siempre se trata de modelos ARMA cuando se habla del método Box-Jenkins

Método Box-Jenkins (cont.):

• En resumen, el método Box-Jenkins tiene como objetivo encontrar un (o varios) modelo(s) *simple(s)*, es decir, un modelo con pocos parámetros ("principle of parsimony") y adecuado, y el método tiene 4 pasos:

1. *Identificación*. Consiste en transformar los datos, si es necesario, para que la hipótesis de estacionariedad sea adecuada, y en elegir los ordenes p,q

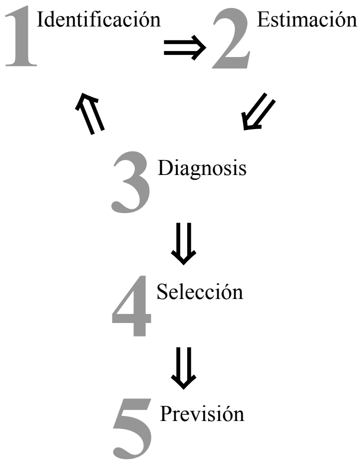
2. *Estimación*. Consiste en estimar el modelo ARMA(p, q)

3. *Diagnosis*. Consiste en comprobar que las propiedades empíricas corresponden a las hipótesis del modelo

4. *Predicción*. Utilizar el modelo para predecir

Método Box-Jenkins (cont.):

- Versión "moderna" del método Box-Jenkins:



Identificación:

• Muchas veces, aplicando el operador diferencia se obtiene una serie estacionaria

• Terminología: Una serie $\{y_t\}$ es *integrada* ("integrated") de orden d si $\Delta^d y_t$ es una serie estacionaria. También se dice que d es el *orden de integración* ("order of integration")

• Notación: $y_t \sim I(d)$ y ARIMA(p, d, q), donde $d \geq 0$

• Si d no es un número entero, por ejemplo, si $d = 0.6$ o si $d = 2.6$, decimos que el orden d es fraccional ("fractional integration")

• *Nota:* En este curso no vamos a ver integración fraccional, siempre vamos a tratar las series como integradas de orden entero, es decir, de orden $0, 1, 2, \ldots$

Identificación (cont.):

- Ejemplos:

  $\longrightarrow$ Si y_t no es estacionaria, pero Δy_t lo es, entonces escribimos $y_t \sim I(1)$

  $\longrightarrow$ Si ni y_t, ni Δy_t son estacionarias, pero $\Delta^2 y_t$ lo es, entonces escribimos $y_t \sim I(2)$

  $\vdots$

  etc.

  $\longrightarrow$ Si y_t ya es estacionaria, entonces escribimos $y_t \sim I(0)$

Identificación (cont.):

• Como encontramos el orden de integración adecuado?

• Para apoyar o rechazar una hipótesis de orden $d = 1$, por ejemplo, podemos utilizar varios tipos de información:

(a) Inspección visual de gráficos

(b) Propiedades y tests estadísticos

(c) Sentido común

(d) Teoría

• *Nota*: Hay casos en los que los investigadores no están de acuerdo. Por ejemplo, hay casos en que no están de acuerdo de si $p_t \sim I(1)$ o si $p_t \sim I(2)$, donde p_t denota un índice de precios en logaritmos

Identificación (cont.):

(a) Inspección visual de gráficos:

$\longrightarrow$ Es la serie creciente durante toda la muestra?

$\longrightarrow$ Oscila la serie alrededor de un valor constante?

$\longrightarrow$ Es la magnitud de la oscilación constante?

$\longrightarrow$ Hay puntos de cambio estructurales?

Identificación (cont.):

(b) Propiedades y tests estadísticos

$\longrightarrow$ Los tests estadísticos de estacionariedad los vamos a ver más tarde (Tema VII; contrastes de raíz unitaria)

$\longrightarrow$ Qué son las propiedades de la FAC ("ACF")?

$\longrightarrow$ Qué son las propiedades de la FACP ("PACF")?

Identificación (cont.):

- *Nota*: Aunque la FAC y la FACP pueden ser útiles para determinar el orden de integración, los usamos sobre todo para determinar p y q

- Recordamos: La FAC es la serie $\{\rho_k\}$, donde $\rho_k = \frac{\gamma_k}{\gamma_0}$

- Para estimar ρ_k utilizamos el estimador

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

donde

$$\begin{aligned} \hat{\gamma}_k &= \frac{1}{T} \sum_{t=k+1}^{T} (y_t - \hat{\mu})(y_{t-k} - \hat{\mu}) \qquad k = 0, 1, \ldots, T-1 \\ \hat{\mu} &= \frac{1}{T} \sum_{t=1}^{T} y_t \end{aligned}$$

Identificación (cont.):

- Prueba de la Q:

$$H_0 : \rho_i = 0 \text{ para } i = 1, \ldots, k$$

$$H_1 : \text{al menos un } \rho_i \neq 0$$

$$\text{Box-Pierce} : Q = T \sum_{i=1}^{k} \hat{\rho}_i \;\sim\; \chi^2(k)$$

$$\text{Ljung-Box} : Q = T(T+2) \sum_{i=1}^{k} \frac{\hat{\rho}_i^2}{T-k} \;\sim\; \chi^2(k)$$

- *Nota*: La prueba de Ljung-Box es una modificación de Box-Pierce y mejor en muestras pequeñas

Identificación (cont.):

- El caso MA(q). Si, aproximadamente, $\{\epsilon_t\} \sim N(0, \sigma^2)$:

$$Var(\hat{\rho}_k) = \frac{1}{T}(1 + 2\sum_{j=1}^{k-1} \hat{\rho}_k^2), \text{ para todo } k > q$$

$\rightarrow$ Rechazamos la hipótesis nula $\rho_k = 0$ si $|\hat{\rho}_k| > 1.96 \cdot \sqrt{Var(\hat{\rho}_k)}$. Recuerda: 5% $\Rightarrow$ 1.96 desviaciones típicas

- Recordamos:

$\rightarrow$ AR(p): ρ_k decae (por ejemplo, para un AR(1) tenemos que $\rho_k = \phi_1^k$)

$\rightarrow$ MA(q): $\rho_k = 0$ para todo $k > q$

Identificación (cont.):

• **Definición: Función de autocorrelación parcial (FACP)**. La autocorrelación parcial entre y_t y y_{t-k} es la correlación entre y_t y y_{t-k} menos la parte explicada por $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$:

$$\rho_k^* = Corr[y_t - E(y_t|y_{t-1}, \ldots, y_{t-k+1}), y_{t-k}]$$

donde $E(y_t|y_{t-1}, \ldots, y_{t-k+1})$ es el predictor de mínimos cuadrados de y_t respecto a $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$, y donde $y_t - E(y_t|y_{t-1}, \ldots, y_{t-k+1})$ es el residuo

• La FACP puede ser útil para eligir el orden p

• Por ejemplo, si estimamos un AR(1) pero una estructura AR(2) es mas adecuada, entonces la FACP nos lo indicaría

Identificación (cont.):

- La regresión lineal

$$y_t = \hat{b}_0 + \hat{b}_1 y_{t-1} + \hat{b}_2 y_{t-2} + \cdots + \hat{b}_k y_{t-k} + \hat{e}_t,$$

estimada con Mínimos Cuadrados Ordinarios (MCO), nos da una estimación de ρ_k^*: $\hat{b}_k = \hat{\rho}_k^*$

- Si, aproximadamente, $\hat{\rho}_k^* \sim N(0, \frac{1}{T})$:

H_0: $\rho_k^* = 0$
H_1: $\rho_k^* \neq 0$

$\rightarrow$ Rechazamos H_0 si $|\hat{\rho}_k^*| > \frac{1.96}{\sqrt{T}}$
Recuerda: 5% $\Rightarrow$ 1.96 desviaciones típicas

Estimación:

• Normalmente los modelos ARMA se estiman con Mínimos Cuadrados No Lineales (MCNL, "NLS") y Máxima Veromisilitud (MV, "MLE")

• Para ello se usan métodos numéricos

• Un criterio, entre otros, de éxito: Convergencia

• EViews utiliza MCNL: Se dice "Convergence achieved after x iterations", entonces hay convergencia

• Si se dice "Convergence not achieved after x iterations", entonces no hay convergencia

$\rightarrow$ Cambiar criterios de convergencia; cambiar valores iniciales; cambiar orden p y q

Diagnosis:

- Objetivo: Determinar si el modelo ARMA es adecuado
- Es decir, comprobar que las hipótesis del modelo son adecuadas
- Es $\{\epsilon_t\}$ ruido blanco?

  $\rightarrow$ Comprueba que $Cov(\epsilon_t, \epsilon_{t-k}) = 0$: FAC de $\{\hat{\epsilon}_t\}$
  $\rightarrow$ Comprueba que $Var(\epsilon_t)$ es constante: FAC de $\{\hat{\epsilon}_t^2\}$
  $\rightarrow$ Si suponemos ruido blanco Gaussiano: Test de normalidad (Jarque-Bera)

- Son los parámetros estables?

  $\rightarrow$ Tests de Chow, Test de Ramsey ("RESET")

Diagnosis (cont.):

• Prueba de la Q (Ljung-Box) aplicado a los residuos de un modelo ARMA(p,q):

$$H_0 : Corr(\epsilon_t, \epsilon_{t-i}) = 0 \text{ para } i = p+q, \ldots, k$$

$$H_1 : \text{al menos un } Corr(\epsilon_t, \epsilon_{t-i}) \neq 0$$

$$Q = T(T+2) \sum_{i=p+q}^{k} \frac{\hat{\rho}_i^2}{T-k} \ \sim \ \chi^2(k-p-q-1)$$

• Prueba de Durbin-Watson: Valor igual a 2 $\Rightarrow Cov(\epsilon_t, \epsilon_{t-1}) = 0$

Selección:

• Objetivo: Seleccionar un modelo (o algunos modelos) simple(s) entre varios candidatos adecuados

• Algunos métodos de selección son:

1. Elegir el modelo o los modelos con menor "Akaike Information Criterion" ("AIC") o "Schwarz Bayesian Criterion" ("SBC")
2. Modelación de lo simple a lo general ("simple to general")
3. Modelación de lo general a lo simple ("general to simple")
4. Combinación de 1., 2. y 3.

Selección (cont.):

• Número de parámetros $\uparrow$, $\Rightarrow$ suma de errores cuadrados $\downarrow$ (mejor ajuste)

• Si $s = 1 + p + q$ (número de parámetros), busca modelo con menor:

$\rightarrow$ "Akaike Information Criterion" (AIC): $\ln(\frac{1}{T}\sum_t \hat{\epsilon}_t^2) + 2 \cdot s$

$\rightarrow$ "Schwarz Bayesian Criterion" (SBC): $\ln(\frac{1}{T}\sum_t \hat{\epsilon}_t^2) + \ln(T) \cdot s$

• Es decir, busca modelo con menor AIC o SBC

• SBC penaliza más el número de parámetros que AIC (porque $\ln T > 2$ cuando $T > 7$)

Selección (cont.):

- *Nota*: EViews utiliza versiones de AIC y SBC un poco diferentes

- Si l denota el logaritmo de la función de verosimilitud Gaussiana y s es el número de parámetros (ejemplo: en el caso ARMA(p,q) tenemos que $s = 1 + p + q + 1$) entonces:

    $\rightarrow$ EViews AIC: $-2(\frac{l}{T}) + 2(\frac{s}{T})$

    $\rightarrow$ EViews SBC: $-2(\frac{l}{T}) + \frac{s \ln T}{T}$

- Busca modelo con menor AIC o SBC

- SBC penaliza más el número de parámetros que AIC (porque $\ln T > 2$ cuando $T > 7$)

Selección (cont.):

• Los métodos 2. y 3., es decir, "de lo simple a lo general" y "de lo general a lo simple", buscan también un modelo simple, pero con parámetros significativos

• Es decir, aunque un modelo con menos términos es mejor que otro con más términos según AIC o SBC, 2) y 3) eligen el modelo con más términos si estos son significativos

• Recuerda que—aproximadamente—un parámetro es significativo al 5% si el valor $|t| > 2$, y si los residuos son aproximadamente Gaussianos

Selección (cont.):

- *Ejemplo*. Selección del mejor modelo(s) ARMA de $\Delta \log PIB_t$ utilizando EViews:

	$\hat{\phi}_0$ (p-val)	$\hat{\phi}_1$ (p-val)	$\hat{\theta}_1$ (p-val)	Q (p-val)	Q^2 (p-val)	AIC	SBC
ARMA(0,0)	0.04 (0.00)	-	-	1.81 (0.18)	0.48 (0.49)	-3.45	-3.41
ARMA(1,0)	0.041 (0.00)	0.19 (0.17)	-	0.51 (0.48)	3.56 (0.06)	-3.58	-3.50
ARMA(0,1)	0.04 (0.00)	-	0.25 (0.07)	0.27 (0.61)	1.45 (0.23)	-3.46	-3.38
ARMA(1,1)	0.04 (0.00)	0.14 (0.68)	0.06 (0.87)	1.26 (0.26)	3.42 (0.06)	-3.54	-3.43

Selección (cont.):

- Dado un nivel de significación del 10%:

  → Método 1: Según AIC eliges ARMA(0,1), según SBC eliges ARMA(0,0)

  → Método 2: Eliges ARMA(0,1)

  → Método 3 y 4?

- *Nota*: Utilizando otros tests para el diagnóstico (por ejemplo el test Durbin-Watson) podría cambiar el modelo seleccionado

Previsión:

- Objetivos:

  1. Predecir ("forecast", "predict") una variable económica
  2. Comprobar que "la teoría", es decir el modelo, funciona en la "práctica"
  3. Utilizar la previsión fuera de muestra como un método de selección del modelo

- Teoría de predicción: Tema III

Previsión (cont.):

• Una estrategia sencilla para comprobar si un modelo funciona en la práctica es utilizar una parte de la muestra para la estimación, y la otra parte para la previsión

$\rightarrow$ Si las propiedades del error $\{\epsilon_t\}$ son las mismas dentro y fuera de la muestra, entonces "la teoría funciona en la práctica"

• La previsión fuera de la muestra se puede tambien utilizar como un método de selección de modelo(s)

• Sin embargo, el fallo de previsión ("forecast failure") fuera de muestra puede ocurrir por razones de cambios estructurales, y no necesariamente porque el modelo no es bueno

Referencias:

Box, G. E. and G. M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden-Day. Revised Edition.