Tópicos de Econometría II
Monitoria

Jorge Enrique Muñoz Ayala
Julián David Rojas Aguilar

Bogotá, 19 de septiembre de 2020

Contenido

Hoja de ruta

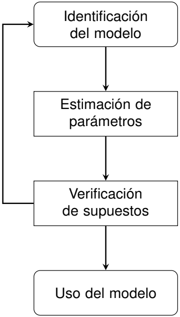
Supuestos:

- La historia se repite.
- No se modelan otras variables explicativas.
- No existe teoría económica que sustente el modelo.
- Se tienen suficientes observaciones.

Objetivos:

- Pronóstico, de corto y mediano plazo.
- Parsimonioso.
- Empleado como modelo *benchmark*.

Figura 1: Metodología de Box y Jenkins (1976).



1. Identificación del modelo: Consta de dos etapas:
   - Determinar una serie estacionaria en función de la serie original.
     - ★ Si la serie no tiene varianza constante, puede aplicarse el logaritmo natural o la transformación Box–Cox para estabilizarla.

Transformación Box–Cox

Sea y_t un proceso estocástico tal que la varianza cambia en función del valor esperado o media μ, $\mathbb{V}\mathrm{ar}(y_t) = c \cdot f(\mu_t)$. Es posible encontrar una función $f(\cdot)$ tal que la varianza sea constante. Esto es:

$$f(y_t) = y_t^{(\lambda)} = x_t = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases}$$

λ es estimado a través de máxima verosimilitud.

     - ★ Si la serie tiene tendencia estocástica, se debe aplicar el operador diferencia, definido como $\Delta = (1 - L) = (1 - B)$.
   - Determinar los valores p y q. Para ello, se emplean los criterios de información. En estos, los órdenes de p y q son tales que se minimiza el criterio de información.
2. Estimación de parámetros:
   - Los modelos AR es posible estimarlos a través de OLS.
   - Los modelos MA o ARMA requieren de métodos de estimación no lineales: Máxima verosimilitud o mínimos cuadrados no lineales.

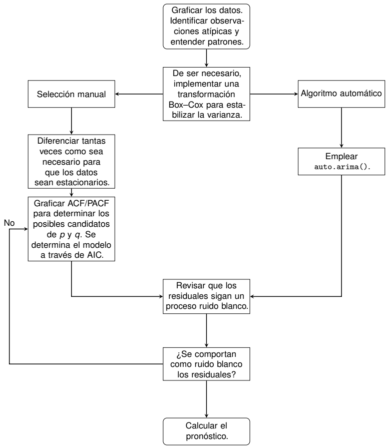
3. Verificación de supuestos:
    1. Prueba de autocorrelación.
    2. Prueba de heterocedasticidad.
    3. Análisis de estabilidad.
    4. Prueba de normalidad.
4. Uso del modelo: Pronóstico de los valores futuros.
    - Supóngase series con varianzas:
        - **Estables**: La transformación logarítmica puede inducir deteriores en la capacidad predictiva del modelo.
        - **Inestables**: La transformación logarítmica demuestra mejora en la capacidad predictiva, por lo que se recomienda trabajar las series en niveles, –teniendo en cuenta que, en presencia de cambios estructurales, esto puede cambiar–.
    - No siempre las series transformadas mediante *Box–Cox* mejoran los pronósticos de las series en nivel.

Típicamente, la estimación de un modelo ARIMA requiere:

- Graficar los datos e identificar observaciones atípicas.
- Si es necesario, transformar los datos para estabilizar la varianza. Esto puede hacerse a través de las transformaciones Box–Cox.
- Si los datos son no–estacionarios, se toma primeras diferencias hasta que lo sean.
- Se revisa la ACF y PACF, para determinar qué órdenes de p y q son apropiados en la búsqueda a través de los criterios de información.
- Se revisa los residuales graficando la ACF y realizando la prueba Pormanteau de autocorrelación. Si no se comportan como ruido blanco, es necesario modificar el modelo.
- Una vez los residuales se comporten como ruido blanco, se calcula el pronóstico.

Nótese que el algoritmo Hyndman–Khandakar solo hace los pasos 2–4, por lo que nosotros mismos deberíamos realizar los demás.

Figura 2: Proceso de estimación general.



Contenido

Hoja de ruta

- La variación estacional es una de las fuentes de ruido sobre las series de datos económicos mensuales o trimestrales.
- El ruido proviene de factores no económicos tales como condiciones climáticas, vacaciones escolares, días feriados, Semana Santa y demás.
- Se descompone la serie en tendencia, ciclo, estacionalidad y un componente no explicado o ruido.
- El componente estacional puede dividirse en una variación regular y no regular, y la literatura se ha concentrado en la primera: las variaciones sistemáticas.
- El impacto de los efectos calendario a nivel sectorial varía dependiendo de la participación de la mano de obra en la producción, por lo que hay un alto grado de heterogeneidad.
- Se debe incluir en las variables regresoras el calendario doméstico al ajuste estacional. Esto puede ser a través del número de días festivos por mes o por el número de días hábiles por mes.

- Supóngase que se tiene:

$$\phi(L) = (1 + \phi_1 L + \cdots + \phi_p L^p)(1 + \Phi_1 L^s + \cdots + \Phi_P L^{s\times P}) \tag{1}$$
$$\delta(L) = (1 - L)^d (1 - L^s)^D \tag{2}$$
$$\theta(L) = (1 + \theta_1 L + \cdots + \theta_q L^Q)(1 + \Theta_1 L^s + \cdots + \Theta_Q L^{s\times Q}) \tag{3}$$

  Donde el polinomio (1) hace referencia al proceso autoregresivo, el (2) al orden de integración y el (3) a las medias móviles. Entonces, el modelo SARIMA viene dado por:

$$\phi(L)\delta(L)y_t = \theta(L)\varepsilon_t \tag{4}$$

- *s* indica el número de observaciones por año, de forma tal que la estructura del SARIMA es $(p, d, q)(P, D, Q)_s$, donde p, P son los órdenes autoregresivos regulares y estacionales, y así sucesivamente.

Serie ajustada estacionalmente: *O, lo que es lo mismo, desestacionalizada*. Una serie desestacionalizada corresponde a la serie original menos su componente estacional.

- Por ejemplo, si $y_t = t_t + s_t + c_t + e_t$ (comportamiento aditivo), entonces la serie desestacionalizada es $y_t^{de} = y_t - s_t = t_t + c_t + e_t$. En el caso multiplicativo, sea $y_t = t_t \times s_t \times c_t \times e_t$, entonces la serie desestacionalizada es $y_t^{de} = y_t / s_t = t_t \times c_t \times e_t$.
- Tanto el componente estacional y, por consiguiente, la serie desestacionalizada, no son observables y existen diferentes métodos para estimarlos.

Definición

"*Trigonometric, Box-Cox, ARMA errors, Trend and Seasonal components*". De Livera et al. (2011)

1. Transformación Box–Cox:

$$f(y_t) = y_t^{(\lambda)} = x_t = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases}$$

2. Modelo:

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + d_t$$

Los componentes son:

- Nivel de corto plazo: $l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$.
- Tendencia de corto plazo: $b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t$. Donde b representa la tendencia de largo plazo.

- Componentes estacionales:

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

- Componente irregular:

$$d_t = \sum_{i=1}^{p} \varphi d_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_i + \varepsilon_t$$

El efecto calendario se estima sobre los errores de pronóstico del modelo TBATS:

1. Se estima el error $\hat{\varepsilon}_t = \hat{y}_t - y_t$.
2. Usando la metodología MCO se estima el modelo:

$$\hat{\varepsilon}_t = \sum_{j=1}^{2} \beta_j cal_{t,j} + w_t$$

   Donde:
   - $cal_{t,1}$: Número de días festivos diferentes a domingo en el mes.
   - $cal_{t,2}$: Variable dicotómica que toma el valor de 1 si la semana santa ocurre en el mes.
3. Finalmente, la serie desestacionalizada y excluyendo efectos calendario es:

$$\hat{y}_t^* = \hat{y}_t - \sum_{i=1}^{M} \hat{s}_{t-m_i}^{(i)} - \sum_{j=1}^{2} \beta_j cal_{t,j}$$

Definición

"*Time Series Regression with ARIMA noise, Missing values and Outliers–Signal Extraction in ARIMA time series*"

$$y_t = x_t + \sum_{i=1}^{n} \alpha_i D_i(B) d_{t,i} + \sum_{j=1}^{m} \beta_j cal_{t,j}$$

Gómez y Maravall (1996); Gómez y Maravall (2001a); Gómez y Maravall (2001b)

Los componentes son:

1. Tendencia y estacionalidad: $x_t \sim \mathrm{ARIMA}$
2. Valores atípicos.

$$\sum_{i=1}^{n} \alpha_i D_i(B) d_{t,i}$$

3. Efectos calendario.

$$\sum_{j=1}^{m} \beta_j cal_{t,j}$$

El proceso ARIMA sobre x_t está dado por:

$$\phi(B)\delta(B)x_t = \theta(B)\varepsilon_{xt}$$

Donde:

$$\phi(L) = (1+\phi_1 L + \cdots + \phi_p L^p)(1+\Phi_1 L^s + \cdots + \Phi_P L^{s\times P})$$

$$\delta(L) = (1-L)^d(1-L^s)^D$$

$$\theta(L) = (1+\theta_1 L + \cdots + \theta_q L^q)(1+\Theta_1 L^s + \cdots + \Theta_Q L^{s\times Q})$$

$$\varepsilon_{xt} \overset{iid}{\sim} \mathcal{N}(0,\sigma_x^2)$$

s indica la frecuencia de los datos.

$$\sum_{i=1}^{n} \alpha_i D_i(B) d_{t,i}$$

Donde:

- $d_{t,i}$: Es una variable dicotómica que indica donde está localizado el valor atípico.
- n: Es el número *outliers*.
- $D_i(B)$: Es un polinomio que refleja el tipo de *outlier* y B es el operador rezago.

$$D_i(B) = \begin{cases} 1 & \text{si el } \textit{outlier} \text{ es aditivo (AO)} \\ \frac{1}{1-0.7B} & \text{si el } \textit{outlier} \text{ es transitorio (TC)} \\ \frac{1}{1-B} & \text{si existe un cambio de nivel (LS)} \end{cases}$$

$$\sum_{j=1}^{m} \beta_j cal_{t,j}$$

Donde:

- $cal_{t,1}$: Número de días festivos diferentes a domingo en el mes.
- $cal_{t,2}$: Variable dicotómica que toma el valor de 1 si la semana santa ocurre en el mes.
- β_j: Son los coeficientes asociados a los efectos calendario.

- Tramo permite modelar los datos atípicos, los efectos calendario y determinar los órdenes del modelo ARIMA para x_t, *en ese orden*.
- El algoritmo impone un límite de hasta dos diferencias regulares y una estacional, es decir, $d \leqslant 2$ y $D \leqslant 1$. Una vez con una serie estacionaria, $\delta(L)x_t$, se determina los órdenes de los polinomios restantes, $\phi(L), \theta(L)$, a través del criterio de información SBC.
- En la etapa Seats, se modela la tendencia y la estacionalidad, dividiendo el proceso x_t en dos, un componente estocástico no correlacionado, w_t y una serie estocástica ajustada estacionalmente, v_t, por lo que $x_t = w_t + v_t$.

Contenido

Hoja de ruta

Sea $\{X_t\}_{t \in T}$ un proceso estacionario con función de autocovarianzas tales que:

$$\sum_{|k|<\infty} |\gamma_k| < \infty$$

Definición: La función de densidad espectral de $\{X_t\}_{t \in T}$, o espectro, es:

$$f(\omega) = \frac{1}{\pi} \sum_{|k|<\infty} \gamma_k e^{-i\omega k}, \quad -\infty < \omega < \infty \tag{5}$$

Donde:

- $e^{-i\omega k} = \cos(\omega) + i \sin(\omega)$ e $i = \sqrt{-1}$.
- Nótese que la condición $\sum_{|k|<\infty} |\gamma_k| < \infty$ implica que la serie en (5) converge absolutamente.

1. La densidad espectral es una función real.
2. El espectro es una función con período 2π y par.
3. $f(\omega) \geqslant 0$ para todo $\omega \in [0, \pi]$.
4. La densidad espectral está relacionada con la función de autocorrelación.

Para cualquier $\omega_1 \in (0, \pi]$:

$$\int_0^{\omega_1} f(\omega) d\omega$$

Se interpreta como la parte de varianza de X_t que se asocia a frecuencias ω que son menores o iguales que ω_1. Por consiguiente, $f(\omega)d\omega$ mide la contribución a la varianza de frecuencias en el rango $(\omega, \omega + d\omega)$.

El periodograma basado en la muestra $X_1,\ldots,X_n$ se define como la función:

$$I(\omega)=\frac{1}{\pi}\sum_{|k|<n}\hat{\gamma}_k e^{-i\omega k},\omega\in[0,\pi]$$

El periodograma se puede expresar como

$$I(\omega)=\frac{1}{\pi}\left(\hat{\gamma}_0+2\sum_{k=1}^{n-1}\hat{\gamma}_k\cos(\omega k)\right),\omega\in[0,\pi]$$

$$\int_0^{\pi} I(\omega)d\omega=\hat{\gamma}_0$$

Referencias

- Box, George y Gwilym Jenkins (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- De Livera, Alysha M, Rob J Hyndman y Ralph D Snyder (2011). «Forecasting time series with complex seasonal patterns using exponential smoothing». En: *Journal of the American Statistical Association* 106.496, páginas 1513-1527.
- Gómez, V. y A. Maravall (1996). *Programs TRAMO (Time series regression with arima noise, missing observations, and outliers) and SEATS (Signal extraction in arima time series). Instructions for the User*. Documento de Trabajo. Ministerio de Hacienda España.
- — (2001a). «Automatic modeling methods for univariate series». En: *A course in time series analysis*. Editado por D. Peña, G. C. Tiao y R.S. Tsay. Wiley Online Library. Capítulo 7, páginas 171-201.
- — (2001b). «Seasonal adjustment and signal extraction in econmic time series». En: *A course in time series analysis*. Editado por D. Pena, G. C. Tiao y R.S. Tsay. Wiley Online Library. Capítulo 8, páginas 202-247.
- Hyndman, Rob y Yeasmin Khandakar (2008). «Automatic Time Series Forecasting: The forecast Package for R». En: *Journal of Statistical Software, Articles* 27.3, páginas 1-22.

- Los criterios de información (IC) seleccionan los modelos a partir de la ponderación entre el ajuste del modelo contra una penalidad como consecuencia de la complejidad del mismo al aumentar el número de parámetros.
- La estimación de más y más parámetros tiende a mejorar el ajuste por dentro de la muestra más de lo que en realidad está sucediendo.
- Considere un conjunto de K mopdelos paramétricos, $\mathscr{M}_k$, donde cada modelo $M_k \in \mathscr{M}_K$ $k = 1, \ldots, K$ requiere de estimar n_k parámetros, β_k. Los criterios de información eligen un modelo (dependiente de k) tal que se minimiza:

$$\mathrm{IC}(k) = -2T^{-1} \ln p\left(\hat{\beta}_k \mid z\right) + h(n_k)\, g(T) \tag{6}$$

De entre todos los modelos, $k = 1, \ldots, K$, donde:
  - $p\left(\hat{\beta}_k \mid z\right)$ es la probabilidad de observar los datos dados los parámetros estimados.
  - $h(n_k)$ es una función creciente de la complejidad del modelo, n_k.
  - $g(T)$ es una función decreciente del tamaño de la muestra, T.

Así, $h(n_k)\, g(T)$ penaliza la estimación por el incremento de parámetros (mayor n_k), particularmente en muestras pequeñas. Las formas funcionales de h y g son las que varían entre criterios.

- Considere un conjunto de modelos de regresión lineal que difieren en los predictores seleccionados, x_{kt} :

$$y_{t+1} = \beta_k' x_{kt} + \varepsilon_{kt+1}, \quad \varepsilon_{kt+1} \sim \mathrm{N}\left(0, \sigma_k^2\right)$$

Sea $\hat{\sigma}_k^2$ el estimador por ML de la regresión con n_k variables explicativas, es posible calcular, para los K modelos:

$$\begin{aligned}
\ln p_Z\left(Z_T\right) &= -\frac{T}{2}\ln\left(2\pi\hat{\sigma}_k^2\right) - \frac{1}{2\hat{\sigma}_k^2}\sum_{t=0}^{T-1}\hat{\varepsilon}_{kt+1}^2 \\
&= -\frac{T}{2}\ln\left(2\pi\hat{\sigma}_k^2\right) - \frac{T\hat{\sigma}_k^2}{2\hat{\sigma}_k^2} \\
&= -\frac{T}{2}\ln\left(2\pi\hat{\sigma}_k^2\right) - \frac{T}{2}
\end{aligned}$$

Donde: $Z_T = \left(y_{t+1}, x_{kt}\right), t = 0, \ldots, T-1$. Así, se concluye que:

$$-2T^{-1}\ln p\left(\hat{\beta}_k \mid Z_T\right) = \ln\hat{\sigma}_k^2 + C$$

Siendo C una constante que no varía entre modelos y, por ello, puede ser omitida.

$$\begin{aligned}\mathrm{AIC}(k) &= \ln \hat{\sigma}_k^2 + \frac{2n_k}{T}\\ \mathrm{BIC}(k) &= \ln \hat{\sigma}_k^2 + \frac{n_k \ln(T)}{T}\\ \mathrm{C}_p(k) &= \frac{\mathrm{SSR}_k}{\hat{\sigma}^2} - T + 2n_k\\ \mathrm{HQ}(k) &= \ln \hat{\sigma}_k^2 + \frac{2k \ln(\ln T)}{T}\end{aligned}$$

- Las diferencias entre los distintos IC son:
  - BIC: Selecciona el modelo con la mayor distribución de probabilidad *posterior* dada la información. Suele recomendarse para elegir el número de rezagos en un proceso ARMA.
  - AIC: Busca minimizar la distancia Kullback–Leibler entre la densidad del modelo candidato y la densidad del verdadero pero desconocido modelo poblacional.
  - Mallows: Es un estimador del $\mathbb{E}(\mathrm{MSE})$ aplicado cuando variables relevantes han sido omitidas.
  - Hannan–Quinn: Determina el orden de un modelo autoregresivo lineal. Permite evadir el problema de *overfitting* y su penalidad suele ubicarse intermedio al criterio AIC y BIC.

Algoritmo de Hyndman–Khandakar para la modelación automática

La función `auto.arima()` en R emplea una variación del algoritmo Hyndman–Khandakar, desarrollado en Hyndman y Khandakar (2008). Este combina pruebas de raíz unitaria y minimización del AIC para estimar un modelo ARIMA. Los posibles argumentos de `auto.arima()` permiten variar el algoritmo.

1. El número de diferencias d es determinado a través de la evaluación repetida de la prueba KPSS, donde $0 \leqslant d \leqslant 2$.
2. Los valores de p y q son tales que se minimiza el AIC después de la diferenciación de los datos d veces. Para ello, el algoritmo, en lugar de emplear todas las posibles combinaciones de p y q, hace una búsqueda paso a paso, a saber:
   a. Cuatro modelos iniciales son ajustados:
      - $\text{ARIMA}(0, d, 0)$.
      - $\text{ARIMA}(1, d, 0)$.
      - $\text{ARIMA}(0, d, 1)$.
      - $\text{ARIMA}(2, d, 2)$.

      Una constante es incluida a menos que $d = 2$. Si $d \leqslant 1$, un modelo adicional es estimado:

      - $\text{ARIMA}(0, d, 0)$, sin constante.

- (b) El mejor modelo (aquel con el menor AIC) ajustado en el paso 2a. Este será seleccionado como el "modelo actual".
- (c) Se considera variaciones en el modelo actual, esto es:
  - ★ Estimar p y/o q distintos con ± 1.
  - ★ Incluir o excluir la constante del modelo actual.

  El mejor modelo entre el actual y sus variaciones es escogido como el nuevo mejor modelo.
- (d) Repetir el paso 2c hasta que no pueda hallarse un AIC menor.

El procedimiento por defecto usa una búsqueda con aproximaciones para mayor rapidez. Sin embargo, esto puede ser evitado con `approximation=FALSE`. Así mismo, un mayor espectro de modelos será utilizado si se emplea `stepwise=FALSE`.