Modelado Estocástico

Clase 8

Cuando trabajamos con series de tiempo, lo primero que debemos hacer es verificar el orden de integración de las series de tiempo con las que vamos a trabajar. Si son I(1), las podemos relacionar usando MCO **si están cointegradas**. Si son estacionarias o I(0), las podemos relacionar usando el análisis previo más tradicional, pero teniendo presente potenciales autocorrelaciones.

¿Qué hace exactamente el software cuando le indicamos, por ejemplo, que nuestros errores siguen un proceso AR(1)?

Estimación:

Supongamos que queremos estimar el modelo de regresión simple,

$$y_t = \alpha + \beta x_t + u_t$$

donde los errores siguen un proceso AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

Si estimamos β por MCO obtenemos un estimador que es insesgado pero que no es MELI porque no se cumplen las condiciones del Teorema de Gauss Markov. Ahora bien, si nosotros pudiéramos transformar este modelo de modo tal que obtener errores en el modelo transformado que satisfacen las condiciones de este teorema, entonces podríamos usar Mínimos Cuadrados Ordinarios (MCO) en el modelo transformado.

Notando que:

$y_t = \alpha + \beta x_t + u_t$ (1)

$y_{t-1} = \alpha + \beta x_{t-1} + u_{t-1}$ (2) = (1) rezagado

$\rho y_{t-1} = \rho\alpha + \rho\beta x_{t-1} + \rho u_{t-1}$ (3) = (2) * ρ

Restando de la ecuación (1) la ecuación (3) llegamos a que:

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + \varepsilon_t$$

que podemos escribir de la siguiente manera:

$$y_t{}^* = \alpha^* + \beta x_t{}^* + \varepsilon_t \qquad (4)$$

donde $y_t{}^* = y_t - \rho y_{t-1}$, $\alpha^* = \alpha(1-\rho)$, $x_t{}^* = x_t - \rho x_{t-1}$

Note que los errores de (4) son ruido blanco, de manera que si usamos MCO para estimar los parámetros en (4) nuestros estimadores serán MELI (mejor estimador lineal insesgado). El estimador de β usando MCO en (4) se llama estimador por el Método de Mínimos Cuadrados Generalizados (MCG o GLS en inglés) del modelo original (1).

.

Estimación: procesos univariados

Metodología de Box-Jenkins:

(1) Identificación: transformar la serie de modo que sea débilmente estacionaria y hacer un *guess* inicial de p y q; estamos estimando un ARMA(p,q). Para este *guess* nos basamos en el autocorrelograma y el autocorrelograma parcial.

(2) Estimación: estimamos p y q

(3) Hacemos tests de diagnósticos sobre los residuos para ver si son ruido blanco

Ejercicio: Precio Internacional del Trigo

Fuente: World Bank Commodity Price Data (The Pink Sheet)
https://www.worldbank.org/en/research/commodity-markets

Vamos a usar el período de 1980m1 a 2018m12 (T=468):

Vamos a analizar $\Delta \ln\left(S_t\right) = \ln\left(S_t\right) - \ln\left(S_{t-1}\right)$ para la serie del precio del trigo SRW (dólares por tonelada). Una primera cuestión es ver si la serie es estacionaria o no. Para aplicar la metodología Box-Jenkins necesitamos que sea estacionaria.

Entonces, antes de empezar con Box Jenkins, debemos correr algún test de raíces unitarias (por ejemplo, ADF) sobre la serie, ya que si fuera I(1) debemos tomar diferencias. Necesitamos trabajar con una serie de tiempo ESTACIONARIA.

Una vez verificado que dlws es estacionaria $(\Delta \ln(S_t))$, podemos comenzar aplicando la metodología. Si ocurre que la serie de tiempo con la que vamos a trabajar fuera I(1), debemos tomar diferencias una vez para que la serie en diferencias sea I(0) y continuamos el análisis con la serie en diferencias (para hallar *p* y *q*).

En ARIMA(p,d,q), el “p” denota la parte autorregresiva, el “q” la parte de medias móviles, mientras que la “d” indica el orden de integración de la serie, es decir, la cantidad de veces que debemos tomar diferencias para que la serie resultante sea I(0).

Si la serie de la que realizaremos el análisis es I(0), entonces d=0 y tenemos un proceso ARMA(p,q).

“q” se determina con el gráfico de la función de ACF en un MA puro, mientras que “p” se determina con el gráfico de la función de PACF en un AR puro

Seleccionamos: ARMA(1,1) , ARMA(1,8)

Paso 2: Estimación

Vamos a estimar los modelos que encontramos como *guess* en el paso 1 y vamos a seleccionar el más apropiado siguiendo diferentes criterios de selección:

a) Akaike (AIC) y Bayesian (BIC): buscamos el menor valor
b) Máxima Verosimilitud: buscamos el mayor valor

```
ARIMA regression

Sample:  1980m2 - 2018m12                       Number of obs     =        467
                                                Wald chi2(2)      =      25.41
Log likelihood =  619.5136                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
        dlws |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
dlws         |
       _cons |   .0005513   .0035557     0.16   0.877    -.0064178    .0075204
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.0303346   .2707296    -0.11   0.911    -.5609549    .5002858
             |
          ma |
         L1. |   .2126807   .2644966     0.80   0.421    -.3057232    .7310846
-------------+----------------------------------------------------------------
      /sigma |    .064213   .0015653    41.02   0.000     .0611451    .0672809
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs   ll(null)   ll(model)      df          AIC         BIC
-------------+---------------------------------------------------------------
           . |        467          .    619.5136       4    -1231.027   -1214.442
-----------------------------------------------------------------------------
               Note: N=Obs used in calculating BIC; see [R] BIC note.
```

Criterio	ARMA(1,1)	ARMA(1,8)	Mejor
Log Likelihood	619.5136	623.99	ARMA(1,8) (>)
Akaike (AIC)	-1231.03	-1237.97	ARMA(1,8) (<)
BIC	-1214.44	-1217.24	ARMA(1,8) (<)

```
arima dlws, arima(1,0,1)
arima dlws, ar(1) ma(7,8)
estat ic
```

Paso 3: Diagnóstico:

(a) Residuos ruido blanco: Portmanteau Test

(b) Que sea estacionario: inversa de raíces dentro del círculo unitario

(c) Que sea invertible

(a) Realizamos un test de diagnóstico que consiste en que los residuos obtenidos en el modelo estimado seleccionado en el Paso 2 satisfagan que son ruido blanco. El test que realizaremos se llama Test de Portmanteau. Bajo la hipótesis nula la serie en cuestión que estemos testeando es ruido blanco, de modo que ahora buscamos no rechazar H0.

```
arima dlws, ar(1) ma(7,8)
predict resid1, residuals
wntestq resid1
```

(b) , (c) Para conocer las raíces inversas, usamos

```
estat aroots
```

Stata muestra los valores y un gráfico.

Si los tres pasos del diagnóstico se satisfacen, entonces podemos decir que hemos encontrado una representación de la serie de tiempo. Si no, debemos elegir otro de los guesses del paso (2) y hacer este mismo análisis.

Regresión Espuria

Si en una regresión algunas o todas las variables son I(1), los estadísticos que utilizamos usualmente pueden o no ser válidos. Cuando los regresores (es decir, la variables) son I(1) y no están cointegrados, los estadísticos que solemos usar ya no son válidos y a esta regresión se la conoce como regresión espuria. En este caso, no existe una combinación lineal de las variables que sea I(0). (Granger & Newbold 1974)

Sea $\boldsymbol{y_t} = (y_{1t} \quad \cdots \quad y_{kt})'$ un vector de *k* x 1 variables que no están cointegradas. Usando la partición $\boldsymbol{y_t} = (y_{1t}, \quad \boldsymbol{y_{2t}}')'$, considere la regresión por Mínimos Cuadrados Ordinarios de y_{1t} en $\boldsymbol{y_{2t}}$, dada por la ecuación estimada $y_{1t} = \widehat{\boldsymbol{\beta_2}}' \boldsymbol{y_{2t}} + e_t$. Como y_{1t} no está cointegrado con $\boldsymbol{y_{2t}}$, el valor verdadero de $\boldsymbol{\beta_2}$ es cero. Se puede demostrar (Phillips 1986) que, en este caso de regresión espuria (es decir, los residuos e_t son I(1)),

- $\widehat{\boldsymbol{\beta_2}}$ no converge en probabilidad a cero y en cambio, converge en distribución a una variable aleatoria que **no** es normal y que no necesariamente tiene media igual a cero.
- Los estadísticos t por MCO para testear si los elementos de $\boldsymbol{\beta_2}$ son iguales a cero divergen a $\pm\infty$ cuando $T \rightarrow \infty$.
- El R^2 de la regresión por MCO converge a 1 cuando $T \rightarrow \infty$.

- La regresión con variables que son I(1) solo tiene sentido cuando las variables están cointegradas.
  https://www.tylervigen.com/spurious-correlations

Cointegración

Sea $\boldsymbol{y_t} = (y_{1t} \quad \cdots \quad y_{kt})'$ un vector de dimensión (*k* x 1) de series de tiempo I(1). Decimos que $\boldsymbol{y_t}$ está cointegrado si existe un vector $\boldsymbol{\beta}$ de dimensión (*k* x 1), o sea, $\boldsymbol{\beta} = (\beta_1 \quad \cdots \quad \beta_k)'$ tal que $\boldsymbol{\beta}'\boldsymbol{y_t}$ sea I(0). Es decir, las series de tiempo no estacionarias en $\boldsymbol{y_t}$ están cointegradas si existe una combinación lineal entre ellas que es I(0).

La combinación lineal $\boldsymbol{\beta}'\boldsymbol{y_t}$ en general es referida como una relación de largo plazo dada por alguna relación de equilibrio entre las variables, motivada por la teoría económica o financiera.

Intuitivamente, las variables I(1) no pueden alejarse demasiado de la relación de equilibrio (en el tiempo) porque van a haber fuerzas que operan para que se vuelva al equilibrio.

Ejemplos de Cointegración:

Modelos de demanda de dinero: implican cointegración entre dinero, nivel de ingreso, precios y tasas de interés.

Teoría de la PPA: implica cointegración entre el tipo de cambio nominal, el nivel de precios foráneo y el nivel de precios doméstico.

Modelos de Ingreso Permanente: implica una relación de cointegración entre el nivel de ingreso y el consumo

Paridad de tasas de interés cubierta: implica una relación de cointegración entre la tasa de interés doméstica, la tasa de interés foránea, el tipo de cambio spot y el tipo de cambio futuro.

La ecuación de Fisher implica una relación de cointegración entre tasa de interés y tasa de inflación.

Estos son algunos ejemplos de relaciones de largo plazo entre variables.

En finanzas, cointegración puede analizarse con datos de alta frecuencia o de baja frecuencia. En general, cointegración con datos de alta frecuencia está motivada por relaciones de arbitraje.

Por ejemplo, por la ley de un único precio, el precio de la acción en un mercado debería ser igual al precio de esa misma acción que cotiza en otro mercado. Esto sugiere una relación de cointegración entre precios de un mismo activo financiero que cotizan en diferentes mercados.

Referencias

- Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, *2*(2), 111-120.
- Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, *90*(429), 268-281.
- Phillips, P. C. (1986). Understanding spurious regressions in econometrics. *Journal of econometrics*, *33*(3), 311-340.
- Schwert, G. W. (2002). Tests for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics*, *20*(1), 5-17.