Universidad de San Andrés
Maestría en Inteligencia Artificial
Modelado Estocástico

Ejercitación 8

Estos ejercicios están planteados en código Stata y pueden responderlos en código Stata o Python, como usted prefiera.

Ejercicio 1 - Simulación de Regresión Espuria.

Consideremos el sistema bivariado $\boldsymbol{y_t} = (y_{1t}, y_{2t})'$ donde:

$$y_{1t} = y_{1t-1} + u_t, \text{ donde } u_t \sim N(0,1)$$

$$y_{2t} = y_{2t-1} + v_t, \text{ donde } v_t \sim N(0,1)$$

Acá, claramente y_{1t} e y_{2t} son series no estacionarias (son caminos aleatorios, es decir, *random walks*).

a) Usando Stata, genere 250 números aleatorios para las variables u_t y para v_t. Ayuda: para generar 250 realizaciones de una variable aleatoria normal estándar en Stata debe usar los comandos:
```
set obs 250
set seed 1
gen u=rnormal()
```
b) Genere las variables y_{1t} e y_{2t} (son las sumas acumuladas de las realizaciones de las variables aleatorias generadas). Note que en (a) usted generó dos variables aleatorias normales independientes entre sí. Suponga un valor inicial de ambas variables igual a cero. Ayuda: para generar la suma acumulada use el comando:
```
gen y1=sum(u)
```
c) Genere una variable tiempo con `tsset` (por ejemplo, usando "_n") y grafique las series y_{1t} e y_{2t} .
d) Corra una regresión de y_{2t} en y_{1t} con constante y reporte el estadístico t que obtiene para el estimador de la pendiente. Observe que es un valor altísimo, cuando en definitiva usted regresó dos variables que fueron generadas mediante la acumulación de variables aleatorias normales, independientes entre sí.
e) Reporte el estadístico de Durbin Watson (autocorrelación de orden 1). ¿Los residuos de esta regresión suman cero?

f) Inspeccione la función de autocorrelación y la de autocorrelación parcial de los residuos y corra un test de raíces unitarias sobre los residuos de esta regresión. Esto es lo que se conoce como regresión espuria, a partir de un paper de Granger y Newbold (1974)[1]. Como las variables y_{1t} e y_{2t} fueron generadas a partir de la acumulación de realizaciones de dos variables aleatorias normales independientes, el valor verdadero del parámetro de la pendiente (el beta) debería ser cero, ya que y_{1t} e y_{2t} no tienen relación alguna. Sin embargo, al regresarlas, uno obtiene que el estimador de la pendiente es estadísticamente significativo, al observar un estadístico "t" alto en valor absoluto, y rechazaría la hipótesis nula que la pendiente beta es cero. Esto es la regresión espuria. Estamos regresando dos series NO estacionarias, que NO están cointegradas (los residuos de la regresión no son estacionarios). La regresión que acaba de correr no sirve y desde el punto de vista estadístico sería incorrecto sacar conclusiones a partir de ella.

Ejercicio 2: Modelo de Corrección de Errores (VECM) Bivariado. En este ejercicio vamos a usar el ejemplo II.5.10 de Carol Alexander de UK *rates*, con las tasas de rendimiento a 1 y 12 meses.

a) Grafique las series. Verifique que cada una de las tasas de rendimiento son I(1).
b) Utilizando algún criterio de selección, decida qué número de rezagos va a utilizar (comando `varsoc`).
c) Utilizando una probabilidad de error tipo I de 1%, verifique, utilizando el test de la traza, que existe una relación de cointegración. Use la opción `trend(constant)`
d) Estime el VECM con las opciones `trend(constant)` y `lags(2)` y analice la estabilidad del mismo.
e) Verifique que $\boldsymbol{\Pi} = \boldsymbol{\alpha\beta}'$. ¿Qué representa el $\boldsymbol{\alpha}$ y qué representa el beta?
f) ¿Alguna de estas tasas causa en sentido de Granger a la otra?

Ejercicio 3: Modelo de Corrección de Errores (VECM) Multivariado. En este ejercicio vamos a usar el mismo ejemplo II.5.10 de Carol Alexander de UK *rates*, con las tasas de rendimiento a 1, 2, 3, 6, 9 y 12 meses.

a) Grafique las series. Verifique que cada una de las tasas de rendimiento son I(1).
b) Utilizando algún criterio, decida qué número de rezagos va a utilizar (comando `varsoc`).
c) Utilizando una probabilidad de error tipo I de 1%, verifique, utilizando el test de la traza, que existen 4 relaciones de cointegración. Use la opción `trend(constant)`
d) Estime el VECM con 2 rezagos y analice la estabilidad del mismo.

[1] Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, *2*(2), 111-120.

e) Muestre las 4 relaciones de cointegración que obtuvo, es decir, las relaciones de largo plazo.

Ejercicio 4: Modelo VAR

Utilizando la base vista en clase de datos Precios y Dinero, suponga que a usted lo contrata el gobierno y le pide que investigue si una mayor tasa de crecimiento de la base monetaria causa en sentido de Granger una mayor tasa de inflación. Para ello, le pide que usted decida cuántos meses hay que tomar para encontrar causalidad en sentido de Granger, si es que hay. En clase, vimos que usando pocos rezagos no encontrábamos causalidad en sentido de Granger. Es posible que haya que tomar más rezagos.

¿Es posible mostrar que una mayor tasa de crecimiento de la base monetaria causa en sentido de Granger una mayor tasa de inflación? Responda esta pregunta planteando un VAR en Stata usando la misma base de datos subida a la plataforma, y usando todo el período. Si fuera posible mostrar causalidad en sentido de Granger, el VAR debe ser estable y el número de rezagos debe estar elegido siguiendo algún criterio de selección óptima de rezagos.

A los fines de este ejercicio, utilice un nivel de significancia (probabilidad de error tipo I) de 10% en todos los tests que realice.

Ejercicio 5: Modelo de Probabilidad Lineal, Probit y Logit

Este ejercicio se responde con la base de datos MROZ que está subida a la plataforma. Para responder, puede usar Python o Stata o el software que usted desee.

a) Corra una regresión regresión por mínimos cuadrados ordinarios donde la variable dependiente es el logaritmo del salario (lwage) y las independientes son una constante, años de experiencia (exper), años de experiencia al cuadrado (expersq), años de educación (educ), edad en años (age), hijos de 6 años o menos (kidslt6) e hijos entre 6 y 18 años (kidsge6). Muestre la regresión.

b) Testee que los betas que multiplican a las últimas 3 variables explicativas (age, kidslt6 y kidsge6) son simultáneamente iguales a cero contra la alternativa que al menos uno de los betas es distinto de cero. Muestre el estadístico F y el p-value asociado a este test. Muestre que no se rechaza la hipótesis nula.

c) Vuelva a correr la regresión de (a) eliminando las 3 últimas variables que resultaron no significativas.

d) ¿En cuánto aumenta el salario en puntos porcentuales si los años de experiencia aumentaran de 4 a 6?

e) Ahora corra una regresión por Mínimos Cuadrados Ordinarios donde la variable dependiente es inlf ("in labor force", note que toma solo los valores 0 ó 1, donde el valor 1 quiere decir que la persona está dentro de la fuerza laboral) y como variables explicativas tome una constante, non-wife income (nwifeinc), educ, exper, expersq, age, kidslt6 y kidsge6. Muestre la regresión. ¿En cuánto aumenta la probabilidad de estar dentro de la fuerza laboral si los años de educación aumentaran en una unidad (un año)?

f) Ahora corra la misma regresión que en el ítem anterior pero usando Probit. Muestre la regresión obtenida. ¿Puede indicar un valor aproximado de en cuánto aumenta la probabilidad de estar dentro de la fuerza laboral si los años de educación aumentaran en una unidad (un año)?

g) Ahora corra la misma regresión que en el ítem anterior pero usando Logit. Muestre la regresión obtenida. ¿Puede indicar un valor aproximado de en cuánto aumenta la probabilidad de estar dentro de la fuerza laboral si los años de educación aumentaran en una unidad (un año)?

h) ¿Cuál es el "*odds ratio*" (razón de probabilidades) de tener un año más de educación?

Aclaración: la razón de probabilidades se define como $\frac{p}{1-p}$.